MEANING IN ANIMAL COMMUNICATION

Varieties of meaning and their role(s) in explaining communication

A thesis submitted for the degree of Doctor of Philosophy, in Philosophy

The Australian National University

November 2018

Statement

This thesis is solely the work of its author. No part of it has previously been submitted for any degree, or is currently being submitted for any other degree. To the best of my knowledge, all help received in preparing this thesis and all sources used have been duly acknowledged.

D. Kalkman

David Kalkman

November 2018

Acknowledgments

Having thesis supervisors who are always generous with their time and quick to return writing, filled with extensive comments, is one thing. Comradery with said supervisors is another. Thus, I warmly thank Kim Sterelny and Ron Planer for their guidance and friendship (and wine) throughout the completion this thesis.

Various chapters, in one form or another, were presented at conferences, workshops and departmental seminars. I thank audiences at the Australian National University, numerous Sydney-ANU (SANU) Philosophy of Biology workshops, and the 2015 International Society for the History, Philosophy, and Social Studies of Biology (ISHPSSB) conference.

People who've helped me along the way, personally and/or professionally, include: Lisa Cox, Andrea Bambrick, Mark Johnson, Kim Shaw-Williams, Chris Lean, Ed Elliott, Jess Isserow, Matt Spike, Lachlan Walmsley, Stephen Mann, Ben Fraser, Cameron Rouse, John Matthewson, Anton Killin, Heather Browning, Ivan Gonzalez Cabrera, Adrian Currie, Alma Barner, Justin Bruner, Richard Moore, Bronwyn Finnigan, Rachael Brown, Carl Brusse, John Zerilli, Don Nordblom, Erick Llamas, Hezki Symonds, Emma Davies, Ross Pain, Dominie Sophia Dessaix, and Bruno Ippedico.

Last, but not least, my family helped me in numerous, non-trivial, ways while working on this project. For their continued support (and health), I am especially thankful.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

Abstract

Why explain the communicative behaviours of animals by invoking the information/meaning 'transmitted' by signals? Why not explain communication in purely causal/functional terms? This thesis addresses active controversy regarding the nature and role of concepts of information, content and meaning in the scientific explanation of animal communication. I defend the methodology of explaining animal communication by invoking the 'meaning' of signals, and responds to worries raised by sceptics of this methodology in the scientific and philosophical literature. This involves: showing what facts about communication a non-informational methodology leaves unexplained; constructing a well-defined theory of content (or 'natural meaning') for most animal signals; and getting clearer on what cognitive capacities, if any, attributing natural meaning to signals implies for senders and receivers. Second, it weighs into comparative debates on human-nonhuman continuity, arguing that there are, in fact, different notions of meaning applicable to human communication that have different consequences for how continuous key aspects of human communication are with other species.

Contents

Preface

Concepts of information and meaning appear frequently in biological explanations. In the realm of developmental biology, genes are often said to be the 'blueprint' of the organism, which are 'transcribed', 'translated', 'edited', and 'proof read'. In the realm of evolutionary biology, major transitions are credited to improvements in both the amount and accuracy of information capable of being 'transmitted' down the generations. Some have even identified the role of genes in evolution with an 'informational' domain, existing independently of the domain of matter and energy[1].

This explanatory practice has gained philosophical attention. Why characterise the role of only some entities in development or evolution as informational? In the developmental realm, sceptics have argued that characterising development as an ongoing process of instruction clouds our understanding of the causal mechanisms responsible. The charge of genetic determinism has been levelled at those who attribute information to genes, but not to other developmental factors like nutrition or sunlight. In the evolutionary realm, it has even been argued that a focus on inter-generational information transmission promotes an implicit dualist ontology.

When it comes to animal communication[2], attributing information to signals is also commonplace. A vervet vocalisation is said to carry information about the presence of a leopard (Seyfarth et al. 1980). Lizard 'pushups', produced for potential predators, are said to provide information about escape ability (Leal 1999). Similarly, the waggle-dance of a honeybee orientated at 40 degrees from the vertical of the hive is said to carry information about food at some distance along a line 40^0 from the sun's azimuth (von Frisch 1967). In these cases, and many more, animal communication is described not merely in causal and/or functional terms. Senders are sources of information about the world (including their own internal states); signals are imbued with this information; and receivers derive this information from signals.

- Information vs influence -

[1] See Godfrey-Smith & Sterelny (2016) for an overview.
[2] 'Animal communication' will mean nonhuman animal communication unless otherwise specified.

A similar, slightly more recent, philosophical debate about information has taken place here too. Some attack the practice of attributing information (or 'content') to animal signals (Rendall, Owren & Ryan 2009; Owren, Rendall & Ryan 2010), while others defend it (Seyfarth et al 2010; Carazo & Font 2010). Sceptics argue that we should conceive of animal communication in terms of influence as opposed to information. Their motivations are both mechanistic and evolutionary (Stegmann 2013). On mechanisms, sceptics argue that declaring a receiver to have 'derived information' from a signal inflates the cognitive sophistication of response behaviours if all that happens is that a phylogenetically or ontogenetically-fixed response is triggered by the signal. Animal communication is not proximately akin to human communication. As a result, we should not impose concepts drawn from the study of semantics ('content') and pragmatics ('meaning') onto explanations of animal communication. When it comes to evolution, sceptics argue that an informational methodology gets the game theory backwards. Attributing content to signals ignores the degree to which we should expect communication to be sender driven and exploitative. Signals, they insist, are phenotypes of senders not receivers. Again, the alleged result of an informational methodology is that we fail to pay enough attention to how animal communication differs from human communication. Instead of focusing on the exquisite ways evolution tinkers with the physical form of signals to influence receivers in ways favourable to senders, we instead focus on the vaguely-defined 'content' that signals 'carry'. What is this content and how does a signal carry it? Why focus on it instead of the intrinsic physical properties of signals that have actual causal influence?

Proponents of an informational methodology respond that much of this is either confused or question begging. First, we needn't posit that receivers derive information from signals in the way children derive sandwiches from lunchboxes. Signals could 'carry information' in virtue of something as simple as correlating with particular states of the world (Scarantino 2013). Imagine a worker with a 'stop' sign standing a few hundred metres down the street from some road works. When the worker displays the sign to oncoming traffic, this correlates with a truck blocking the road at the location of the works. Similarly, a vervet alarm could carry information about a predator being here now simply by correlating with the presence of a predator. There is no need to posit that the acoustic pattern emitted by the sender vervet is some kind of container carrying a distinct commodity known as 'information content'. Second, they argue that the ones getting the game theory wrong are the information sceptics (Seyfarth et al 2010). Receivers must be gaining fitness benefits from being influenced by signals, or else receivers would evolve or learn to ignore signals, given enough time. Proponents argue that we cannot hope to understand why receivers respond to signals, given the response plasticity granted them by natural selection and/or learning, without positing that receivers derive fitness benefits from signals. 'Information' tells us something about the nature of this benefit.

Thus did the information sceptics rattle their sabres and information proponents dig in their heels. But now that the dust has settled a little, where are we left and what is there still to do? This thesis is broadly pro-informational. I offer some important clarifications to an informational approach in the wake of Stegmann's (2013) anthology on the information vs influence debate. On the evolutionary front, I maintain that receivers possess enough flexibility to make information explanatory. Even if receivers are at some disadvantage, information is needed to account for why receivers respond in species-typical fashion. As mentioned, sceptics object to information for reasons that are not only evolutionary. They argue that an informational methodology inflates the cognitive sophistication of communicative behaviours. Taking this criticism on board, I argue that information proponents must be more precise about how receivers 'derive information' from signals. Proponents argue that because receivers possess response flexibility, they must gain fitness benefits if they are 'listening' (e.g. Seyfarth et al. 2010). What they gain is said to be information, derived from signals (Maynard-Smith & Harper 2003, Searcy & Nowicki 2005). I argue that there are two different ways in which organisms can be said to 'derive' information from signals, corresponding to different degrees of cognitive sophistication, and also to two different kinds of response flexibility: diachronic vs synchronic. By not being clear about this distinction, proponents of information have opened themselves up to inflating the cognitive sophistication of receivers in cognitively-simpler species. On the one hand, information can be derived by receivers from signals in a minimal sense. This is when the cognitive link between a mental representation of the stimulus (in this case the signal), and behaviour, is relatively fixed. In such cases, a signal's information explains why receivers are wired-up to respond as they do, from a historical and population level perspective only. On the other hand, information can be derived by receivers in a cognitively-richer sense. This is when the cognitive link between a mental representation of the signal-stimulus, and behaviour, is more sophisticated: i.e. where the organism itself has more control over its response to the signal. Here, the information derived from signals can feature in more-or-less proximate explanations of why individual receivers respond as they do, in real time. It is only in the second case that some degree of cognitive sophistication is implied (although it falls short of human communication). I look at a prominent pro-information paper where the distinction just outlined is not clear (Seyfarth et al. 2010), and where, as a result, an informational methodology can easily be seen as inflating the cognitive sophistication of many receivers. This will be the primary focus of Chapter 3.

A preliminary issue, though, concerns making the notion of signal-content precise. In addition to the worries presented above, sceptics have also complained that the sense in which signals 'carry information' is not spelled out in enough detail by proponents (Rendall, Owren & Ryan 2009; Owren, Rendall & Ryan 2010). Instead, proponents (especially practicing biologists) make vague gestures towards Shannon's (1948) notion of information, formulated for quite different ends in the field of communications engineering. To this end, Chapter 1 surveys a range of well-articulated accounts of content originating in the philosophy of mind literature, plus some philosophical theories of content that have already been advanced for animal signals. Drawing on the concerns of biologists, I argue that the kind of content attributed to animal signals should makes sense of why receivers respond to signals. This is a requirement that constrains my choice of theory for grounding signal content, of which there are many. To this end, I defend an analysis of content wedding correlation with teleology (Shea 2007). The content of a signal is not simply what the latter correlates with, as argued by, for instance, Skyrms (2010), Scarantino (2013; 2015), and Birch (2014). This is because any signal will correlate with a) too many things, which b) won't necessarily rationalise the receiver's response. Content must be whittled down by considering the selected response of the receiver: more specifically the world-state that historically obtained when the response elicited by the signal was beneficial to the receiver (Millikan 2004). Conversely, though, the content of a signal is not simply the world-state that historically obtained when the elicited response was beneficial to the receiver, as argued by Artiga (2014) and Stegmann (2005a; 2009). It must, in addition, be the world-state the signal correlates with. This is because, without correlation, explaining historical success becomes circular (Godfrey-Smith 1996, Shea 2007).

A third unresolved issue is whether communication, or signalhood, can be uniquely defined from an informational perspective (Scott-Phillips 2008; Carazo & Font 2010; Scarantino 2013). Like in the philosophical literature on organismal development and evolution, the question is this: why do we label some causes in nature as 'informational', but not others? When it comes to the current debate, why do we call the intermediary causal factors in some co-adapted interactions (i.e. communication) but not others (i.e. grooming) 'signals'? The answer, as I will show, can't be that signals, uniquely, correlate with functionally-relevant world states (Scarantino 2013)[3]. This is too liberal. Nor can the answer appeal to the 'arbitrariness' of signals vs non-signals. This is too restrictive. Debate has occurred over whether the notion of arbitrariness might be lent upon to defend the attribution of information to genes as opposed to other developmental factors (Godfrey-Smith 2000, Stegmann

[3] This is so even when content is grounded by teleology as well as correlation, as it will in this thesis.

2005b). However, while arbitrariness may justify the singling out of some causal factors as informational but not others when it comes to organismal development, I argue that it cannot do so when it comes to between-organism communication. This is because many animal signals are not arbitrary. Arbitrariness, at least when it comes to between-organism interactions, does not make some phenotypes but not others 'informational' in nature. This argument will involve offering a counterfactual notion of arbitrariness, as the notion is not well-defined in the current literature, along with showing why some paradigm animal signals do not appear to be all that arbitrary under this notion. The overarching moral here is that while an informational methodology is justified and useful for explaining aspects of animal communication, such communication cannot be uniquely defined in informational terms. This will be the topic of Chapter 2.

- *Natural and non-natural meaning* -

In an influential paper published in 1957, Grice distinguished between what he called 'natural' and 'non-natural' meaning. The former he understood as an entailment relation between states of affairs. Fire naturally means smoke because the presence of fire entails smoke. Dretske (1981) sought to naturalise mental content using this kind of relation. Others have followed this line of thinking but have weakened the required relation: natural meaning obtains whenever there is contingency/correlation between states of affairs. Thus, smoke naturally means fire because it raises the probability of fire, as opposed to making fire certain[4]. Inspired in part by Dretske (1981), Skyrms (2010, p. 1) proclaimed boldly that "all meaning is natural meaning".

Chapters 1-3 operate using a notion of natural meaning, albeit one backed by teleology as well as correlation (Shea 2007). Chapter 4 changes focus and engages with Gricean *non-natural* meaning. This is because, despite bold assertions by Skyrms (2010), not all meaning is natural meaning. At least, not all meaning attributed to animal signals in the literature is natural meaning. By 'non-natural meaning', Grice meant something very different from entailment and/or correlational relations. Exactly what he meant will be fleshed out in Chapter 4, but basically for some behaviour performed by a sender to carry non-natural meaning it needs to be 1) intentional and 2) overtly so. The first criteria states that the behaviour must be motivated by a goal which is robustly pursued by an organism sensitive to means-ends contingencies in real time. The second requirement holds that this intention motivating the sender's behaviour, i.e. the sender's "informative intention" (Sperber & Wilson 1986), is embedded within a more sophisticated intention: that the receiver

[4] There are, after all, smoke machines.

recognise the sender's informative intention. This latter intention is sometimes called the "communicative intention" (Sperber & Wilson 1986).

There is a currently-influential line of post-Gricean research on the topic of language evolution and operation (Sperber & Wilson 1986; 2002, Gomez 1994, Tomasello 1999; 2008, Bloom 2000, Hurford 2007, Csibra 2010, Scott-Phillips 2015, Moore 2016; 2017). It posits that both a) acting with the above intentions, and b) recognising when others are doing so, are necessary for human communication. Human communication requires the expression and recognition of intentions, because what is explicitly uttered (or gestured) is always/often an imprecise guide to what speakers aim to achieve by uttering. This is to say that semantic content underdetermines speaker meaning, where non-natural meaning is envisaged as speaker meaning. Jones curses while overtly looking at his watch for Smith. Does Jones want Smith to believe they're late for the lecture, or that he paid too much for his watch? Without the ability to recognise what speakers want to achieve by uttering (or gesturing), human communication wouldn't succeed because hearers would never 'get the message'. The message (or the speaker's meaning) is the speaker's intention: in this case, that the hearer believe they're late for the lecture. Thus, human communication is not simply the coding and decoding of semantic content (and sometimes also natural meaning). Of course, coding and decoding often occurs. Language is obviously a massively important component in human communication. Plus, we sometimes employ natural information in our communicative actions, such as when I point to the dark rings around my eyes in telling you I'm exhausted. But these 'explicit' sources of information are used in the service of a more fundamental process. This is the process of expressing and recognising speaker meaning. Explicit evidence of speaker meaning is provided, which the hearer must then combine with other information about the context, in order to infer the speaker's intended outcome in uttering/gesturing (i.e. that you believe I'm exhausted). In short, human communication is an exercise in folk psychology or 'mindreading'. Explicit evidence, produced by senders, is used by hearers to infer a representation which is a placeholder for the speaker's intention in communicating. This stands in opposition to communication involving natural meaning only. The latter can take place with zero awareness, on the part of senders or receivers, of one another's motivations.

According to the post-Gricean line, then, the ability to express and recognise intentions explains how humans express and recognise (non-natural) speaker meaning. However, a challenge to the post-Gricean view comes from developmental psychology, and is presented by Moore (2017). Expressing and recognising both informative and communicative intent is standardly thought to require, not just the ability to make inferences about others' mental states, but also the ability to make inferences about others' mental states towards one's own mental states: to think things like *S intends me to believe that S wants me to do P or believe P* (Sperber 2000, Tomasello 2008, Scott-Phillips 2015). However, the present developmental literature on higher-order mindreading provides reasons to think that this is too demanding. The degree of meta-representation required to express and recognise non-natural meaning is plausibly beyond the capabilities of young children, who are nevertheless in the process of learning how to express and recognise speaker meaning. Moreover, it doesn't *seem* like adults engage in such inference in quotidian episodes of communication. In response to this, some Post-Griceans suggest that the required meta-representation goes on implicitly and sub-personally (Sperber & Wilson 2002; Scott-Phillips 2015). Others have attempted to reduce the intentional demands of expressing and recognising Gricean non-natural meaning (Moore 2016; 2017). In Chapter 5, I argue that the intentional (and thus meta-representational) demands of communicating with non-natural meaning can be reduced only so far before the relevant intentions become explanatorily redundant within the post-Gricean picture. As a result, I argue that all proponents of Gricean non-natural meaning must take seriously the idea that humans can implicitly/sub-personally meta-represent intentions. In Chapter 6, I offer a crude cognitive model of how this might occur. The model is massively-modular in nature, because the relevant post-Griceans favour a massively-modular view of the mind. However, before this, in Chapter 5, I present a graded picture of speaker meaning. I argue there are kinds of speaker meaning that are a) non-Gricean, while at the same time b) more sophisticated than natural meaning. They are more sophisticated than natural meaning because they require intentional behaviour on the part of speakers, and intention recognition (i.e. mindreading) on the part of hearers. However, the intentions motivating speakers and recognised by hearers are less recursive than Gricean non-natural meaning. Here I draw on the work of Csibra (2010) and Moore (2017), while parting ways with the latter to some degree.

While Chapters 4 and 5 focus largely on human communication, one of the primary objectives is nevertheless comparative. We must know what humans are doing when communicating before we can judge whether other animals are carrying out the same cognitive tasks, or simpler ones. If humans are doing something simpler than post-Griceans think, this has comparative ramifications. If, as suggested by Moore (2016, 2017), the expression and recognition of non-natural meaning is simpler than post-Griceans think, then perhaps other primates are doing it. Alternatively, if human communication is non-Gricean but still more advanced than the sending and receiving of natural meaning, as I will suggest, then some other primates might, alongside humans, be engaged in a form of communication that many other species, being limited to natural meaning, are not.

Finally, Chapter 6 interrogates the foundation of all models of communication featuring the expression and/or recognition of intentions, whether fully Gricean or not. It does this because of the murky relationship between intentional psychology and cognitive explanation (Bermudez 2005). On some views of this relationship, and on decompositional views of cognitive explanation, citing intentional states when explaining behaviour (i.e. belief-desire pairs) is not doing 'genuine' cognitive science. Instead, intentional psychology merely provides a way to rationalise and predict the behaviours of others, in an abstract sense. There need be no systematic resemblance between, on the one hand, states attributed by intentional explanation, and on the other, descriptions of mental processes at lower levels of analysis (Dennett 1987). In response, and instead of concluding so much the worse for intentional psychology, I conclude so much the worse for views of cognitive explanation that always demand decomposition. Even if intentional psychology is as abstract as Dennett maintains, it nevertheless provides a powerful basis for understanding, intervening on, and coordinating with, the thoughts and behaviours of others. This holds in the case of communication, too. As a result, it is not explanatorily empty to posit that human (and perhaps some nonhuman) communication involves the expression and recognition of intentions.

While Chapter 6 is less explicitly concerned with communication, it is continuous with the overall project in interrogating the explanatory role of varieties of meaning more sophisticated than merely natural meaning: more specifically, the explanatory role of intentions, the expression and recognition of which constitutes the expression and recognition of speaker meaning (whether fully Gricean or not). Because it is worth asking whether nonhuman primate communication goes beyond merely the sending and receiving of natural meaning, it had better be the case that this kind of meaning is not explanatorily empty.

Although this thesis is relatively eclectic in nature, it has some unifying themes. Natural meaning is a modest property of signals. It need not worry sceptics of information in the information vs influence debate. Attributing natural meaning to animal signals, by itself, implies little in the way of cognitive sophistication, and little in the way of continuity with human communication. Indeed, the natural meaning of animal signals should usually be seen as part of a broadly ultimate or "structuring" (Dretske 1994) explanation of communicative behaviour. This stands in contrast to speaker meaning, expressed and recognised in species capable of mindreading. Here, recognising speaker meaning (whether fully Gricean or not) requires cognitive sophistication. It involves inferring what speakers intend to achieve by signalling, even if such intentions are not quite as sophisticated as Grice and/or his followers envisage. We can (and I argue we probably should) deflate the kind of meaning humans paradigmatically traffic in, making it simpler than Gricean non-natural meaning. Nevertheless, expressing and recognising speaker meaning depends on cognitive abilities not required by the sending and receiving of natural meaning, alone.

- A note on terms -

Throughout, I use the following terms interchangeably to refer to that property of signals thought problematic by information sceptics and defended by information proponents: *information*; *content*; *natural meaning*. Chapter 1 argues that the property referred to by these terms is grounded by signal-world correlations along with biological proper function. This property is contrasted in Chapter 4 by 'speaker meaning'. The latter is restricted to communication where speakers are motivated by intentions towards hearers, and where hearers must infer the intentions of speakers in order to understand what is meant. These intentions might be Gricean in nature, or alternatively they might be less recursive. In contrast to natural meaning, speaker meaning features paradigmatically in psychological explanation. Although I distinguish a biological notion of natural meaning from psychological notions of speaker meaning in the thesis, in the conclusion I suggest a way in which the two might be combined, at least when it comes to human (and plausibly some great ape) communication.

* * *

Chapter 1. *Why attribute natural meaning to signals? Why not explain animal communication in causal/functional terms alone? Given we do want to attribute natural meaning to signals, what is the nature of this property? What biological facts ground this kind of meaning? There are several theories of content that could be used, each grounding natural meaning in different kinds of biological facts. Which one is most suitable? What explanatory considerations should constrain a good theory of natural meaning for signals?*

Chapter 1

A Biological Notion of Natural Meaning for Signals

The primary objective of this chapter is to argue for a well-defined theory of content for animal signals. A secondary objective is to justify a desideratum that might constrain a good theory of content for animal signals. This secondary goal is important, because theories of content should be constrained by explanatory concerns (Shea 2007). It is one thing to apply some theory of content or another to structures in the world customarily or intuitively regarded as information-bearing. It is another thing for this content to do explanatory work. Drawing from concerns in the biology literature, I argue that a suitable theory of content for animal signals should make sense of why receivers respond to signals (and hence, also why senders produce them).

This desideratum for a theory of content is justified in Section 1. There I also justify explaining animal communication in informational terms and not merely in causal and/or functional terms, given plausible evolutionary assumptions expanded upon in Chapter 3. Sections 2 and 3 then consider a range of theories of content, each drawn from the philosophical literature, to find one that satisfies my desideratum. To bring some order into what has become a rather burgeoning scene, I represent the available theories of content according to three criteria: (1) whether a theory of content is input-orientated or output-orientated; (2) whether it invokes the notion of selected function; and finally (3) whether correlation plays a content-determining role. I argue that my desideratum is best met by a theory of content that is: (1) output-orientated, and (2) inclusive of a notion of selected function (specifically, the selected function of the receiver's response to a signal). Further, I argue (3) that correlation must play a content-determining role within the theory.

1. The Explanatory Role of Content

> When a male courts a female, do his signals honestly convey his quality relative to other males? Or does he exaggerate his quality in order to win over females that would otherwise choose some other male? When one animal signals aggressively in a contest over a resource, does the signaller honestly covey its likelihood of attack? Or does the signaller exaggerate that likelihood in order to intimidate competitors that

> would otherwise defeat him? The question of reliability versus deceit arises even in interactions that, on the face of things, seem to be predominantly cooperative. When an offspring begs for food from its parents, does it honestly convey its level of need? Or does the offspring exaggerate its need to order to get more food than the parents would otherwise provide? (Searcy & Nowikci 2005, p. 1).

Searcy & Nowicki begin their well-known book by highlighting one of the central questions occupying the biological study of animal communication. This is the question of signal reliability, or 'honesty'. Given the predominant view that selection operates on individuals or their genes, and given, also, that often the fitness interests of senders and receivers diverge to varying degrees (sometimes conflicting entirely), one would expect senders to often signal unreliably or 'dishonestly'. Imagine a male of relatively low genetic quality courting a female. One would expect him to mimic the mating signal of a high-quality male, thus influencing the female to mate with him, where this is sub-optimal for her or her genes. After all, selection operates on sender signalling strategies, if it operates on anything in communication. Surely, though, selection *also* operates on receiver response strategies. As a result, the female will ignore most dishonest signals, creating selection pressure for costly, and thus, honest, signalling (Maynard-Smith 2003, Searcy & Nowicki 2005). Sceptics of information in the biology of animal communication argue that this assumption is too quick. While acknowledging receiver flexibility, they argue we should expect senders to nevertheless have the upper hand (Owren, Randall & Ryan 2010). Senders should produce traits tapping into pre-existing sensory biases in receivers, such that receivers are manipulated. In response, information proponents argue that the selection pressures operating solely on senders cannot fully explain why receivers respond. Manipulation can only go so far in explaining why, and just as importantly, *how*, receivers respond. Moreover, as the sceptics themselves admit, a situation in which receivers are being *strongly* manipulated is unlikely to persist. Chapter 3 will go into these issues in more detail. For now, let us assume, plausibly, that receivers have some adaptive 'choice' in responding to signals. If this is indeed the case, then the following is suggested. Signals must be reliable, at least to a significant degree. If they weren't, then receivers would respond in species-typical fashion less often than they do, and, concomitantly, senders would produce less often than they do. But communication is rife in the animal kingdom.

From this, at least two questions are raised. First: what ensures that signals are, at least often enough, reliable? Second: what, exactly, do receivers gain from responding to reliable signals? Here, I am interested in the second question as opposed to the first[5]. I am interested in how we should characterise what receivers

[5] On the first question, Maynard-Smith & Harper (2003) and Search & Nowicki (2005) present a wide range of mechanisms ensuring signal honesty, or at least enough honesty to keep receivers 'listening'. They share informational assumptions about evolutionary parity just discussed.

gain from responding to reliable (enough) signals. More specifically, what property do reliable signals possess that unreliable signals would not (if the latter were produced), such that we can point to that property in explaining why receivers respond? An intuitive answer to this question is that reliable signals carry 'information', where this information is 'about', or has *content* referring to, relevant states of the world (including what senders might do next). That receivers derive this information/content from signals explains why receivers currently respond: i.e. why they have not evolved to ignore signals. Thus, Searcy and Nowicki (2005, p. 208) write "the only value of a signal to a receiver is as a source of information, and without this value the receivers will not respond to the signal and the signalers will not produce it". This same idea motivates Seyfarth et al. (2010, p. 3), in a paper called *the central importance of information in studies of animal communication*, to assert "the concept of information plays a central role in studies of animal communication". According to Seyfarth et al., we cannot make sense of why receivers respond to signals, given the response plasticity granted to receivers by natural selection (and also learning, it must be noted), unless signals carry information.

Similarly, Wheeler et al. (2011, p. 189) state "the only mechanism from which receivers can benefit by responding [to signals] entails signals which vary consistently with some feature of the environment" According to Wheeler et al., signals "varying consistently" with particular features of the environment is how signals carry information. The particular feature of the environment signals covary with is what signals carry information about. In a similar vein again, Maynard-Smith & Harper (2003, p. 3) assert "if a signal alters the behaviour of others it must, on average, pay the receiver of the signal to behave in a way favourable to the signaller; otherwise receivers would cease to respond". This is because signals (unlike acts of coercion) influence receivers in a way that "depends on the evolved properties of the brain and sense organs of the receiver" (Maynard-Smith & Harper 2003, p. 3). An act of coercion like a push does not rely on the evolved properties of the brain and sense organs of the receiver. The 'recipient' has no choice but to 'comply'. Alternatively, when a Peacock's brilliant display causes a receiver to become sexually receptive, this does rely on the evolved properties of the brain and sense organs of the receiver. The receiver has a genuine 'choice' in whether to comply[6], since the brain and sense organs of the receivers undergo selection. Given receivers have some adaptive 'choice' in whether to respond to signals, and given receivers *are* responding to signals, then "It follows that signals must carry information… that is of interest to the receiver" (Maynard-Smith & Harper 2003, p. 3).

[6] This choice might not be one the receiver makes, herself. Instead, it is one which natural selection makes on her behalf.

In short, the following idea is widespread among information proponents. The idea is that signals carry information, where this information makes sense of the existence of communication. More specifically, there is diachronic plasticity in the receiver's response to stimuli. As a result, receivers must be making a return on being causally influenced by signals. If they weren't, then they wouldn't be responding. But receivers *are* responding. Thus, it is suggested, they must be gaining some adaptive benefit. This adaptive benefit is the information derived from signals. I want to use this thinking as a constraint on a good theory of content for animal signals. Thus, I posit that a theory of content should make sense of why receivers respond to signals (and hence why senders produce them). As we will see in Sections 2 and 3, some theories of content are more equal than others on this score. I should note that, in what follows, my desideratum refers explicitly to making sense solely of why receivers respond to signals. However, I take it that making sense of this goes a long way towards making sense of why senders produce signals, too. I refer explicitly to content having to make sense solely of receiver responses, for reasons of brevity. In doing this, though, I do not want to be understood as neglecting intimate evolutionary connections between receiver responses and sender strategies (Godfrey-Smith 2013).

Before we begin shopping for content, however, a potential objection ought to be considered (Scott-Phillips 2008). The objection goes as follows: talk of information carried by signals is superfluous. It is perfectly possible, and equally enlightening, to describe communication in terms of fitness benefits. Indeed, all 'information' stands in for is the fact that receivers, as well as senders, historically gained a fitness benefit by responding to signals. This idea can be captured perfectly well without saying that signals 'transmit information', and without saying that receivers in some sense 'derive' this 'information'. For example, instead of positing that peahens derive information from brilliant peacock tails, we should say (simply) that peahens derive a fitness benefit from responding in species-typical fashion to brilliant peacock tails. While this approach might seem more parsimonious in dispensing with information, it leaves something important out of the picture. What it leaves out is the *reason* receivers gain fitness benefits from responding to signals (Carazo & Font 2010, Lean 2014).

Imagine that, because of selection, two receivers respond differently to a kind of signal: one runs away and the other prepares for a fight. Imagine the information carried by the signal is that the sender is preparing to attack. So here we have one information content (that the sender is preparing to attack) and two different responses to this content (running away and preparing for a fight). Now, we can ask *why* it pays each receiver to respond in their respective way to the signal. A prima facie plausible answer will be because the signal carried the information that the sender is preparing to attack. Without invoking this information, we can't make sense of why our two receivers benefitted from responding to the signal differently. Thus, information is not just a stand-in for the fact that receivers, as well as senders, commonly benefit from communication. Rather, information helps explain why receivers benefit from responding as they do.

Thus, an appropriate theory of content for animal signals must make sense of why receivers currently respond to signals, given that the fitness interests of senders and receivers often diverge. Moreover, content must be the kind of property that makes sense of the evolved responses of our two receivers above.

2. Shopping for Content

Drawing from the concerns of biologists studying the evolution of animal communication, I've proposed that an informational gloss on communication ought to make sense of why receivers respond to signals. On this picture, 'content' should be thought of as a property possessed by a reliable signal, where this property explains why a receiver influenced by the signal typically gained a fitness benefit in the past by responding. Sections 2 and 3 go into more detail about this kind of property: they advance a theory of content for animal signals delivering on my proposed desideratum of content attribution. Three different kinds of content originating in philosophy of mind will be considered: causal-informational semantics; causal-informational (i.e. input-orientated) teleosemantics; and output-orientated teleosemantics. Although each originated with the project of 'naturalising' the content of mental representation, they are *prima facie* applicable (indeed, perhaps more applicable) to animal signals. I aim to show that a particular variant of output-orientated teleosemantics ('infotel-semantics') goes farthest in assigning the right kind of content to animal signals across a range of cases, including deception. Other theories output contents failing to satisfy the desideratum advanced in Section 1. I start by considering causal-informational semantics.

2.1 Causal-Informational Semantics

A variety of closely related theories of content are all motivated by the same essential idea: that the content of a sign is no more than the natural meaning (Grice 1957) it carries. By 'natural meaning' (and in contrast to 'non-natural meaning'), Grice meant an entailment relation between states of affairs. For instance, the presence of rain naturally means that your clothes hanging out to dry will get wet. The presence of smoke naturally means there is a fire. A child's fever naturally means that the child is sick. Different labels for essentially the same kind of theory include 'informational semantics', 'indicator semantics', and 'causal theories of content[7]'. Informational and indicator theories can be grouped together, while causal theories can be considered a subset of informational/indicator theories. Informational theories are boarder than causal theories. Causal theories require that a sign be caused by what it represents, while informational theories do not. This is an upshot for non-causal informational theories. Informational theories, unlike causal theories, leave room for the possibility of misinformation: either in the form of 'deception' or in the form of mistakes. In what follows I will examine non-causal informational theories of content.

The (synonymous) notions of 'correlation' and 'probability-raising' are central to the kind informational theories of content I will address. By saying that S carries information about X, we are saying that S *correlates* with X. And by saying that S correlates with X, we are saying that S *raises the probability* of X. There are different ways of unpacking the notion of information as 'correlation' (see Stegmann 2014 for a detailed overview). Not all of them will be compatible with the way information is evoked in explanations of animal communication. First off, however, it is vital to note that probability-raising (and hence correlation) is to be cleanly differentiated from the notion of 'coincidence'.

Coincidence is not co-extensive with correlation. Even if it's the case S and X coincide to some degree, S doesn't necessary *correlate* with X unless S raises the probability of X. For instance, assume that S never occurs without X. That is, Pr(S/~X) = 0. Assume also that 30 percent of instances in which X obtains S also obtains. That is, Pr(S/X) = 0.3. In this case, S raises the probability of X, or correlates with X, to the order of 30%. But now assume that S *does* sometimes occur in the absence of X (30% of the time to be exact): Pr(S/~X) = 0.3. In this case, although S and X *coincide*, S does not increase the probability of X. This is because S occurs just as much in the absence of X as it does in the presence of X: Pr(S/X) = Pr(S/~X) = 0.3. Thus, S doesn't carry information about X.

Thus, carrying correlational information involves changing probabilities as opposed to mere coincidence. This already rules out various candidates as the content of some signal. For example, despite leopard calls coinciding very reliably with oxygen being in the atmosphere at normal levels, leopard calls do not raise the probability of normal oxygen levels. This is because the leopard call does not make a normal level of oxygen in the atmosphere (NLOIA) more likely than it would be, independent of the signal: Pr(NLOIA/signal) is no higher than Pr(NLOIA).

So, in order for a sign S to carry information about some world-state X, the former must raise the probability of the latter. However, there is a question regarding the degree of correlation required for S to carry (correlational) information about X.

---

[7] Fodorian causal theories of content, not Kripke-Putnam causal theories.

One might think that the required correlation must be maximally strong (e.g. Dretske 1981). For signal S to carry information about X, on this view, the probability of X given S must be 1. There is a problem with this, though. The objective probability that any event in the biological world obtains given the occurrence of some other event will rarely be 1. Thus, such a maximally strong understanding of correlation will not be useful when it comes to thinking about the content of animal signals.

Scarantino (2015), Skryms (2010) and Shea (2007) have each endorsed an account of correlational information that is less demanding than Dretske's[8]. This weaker notion is a good candidate for grounding the content of animal signals. On this weaker understanding, a signal S carrying information about X need not entail X. Rather, S carries information about X obtaining iff the probability of X given S to be greater than the probability of X in the absence of S. For example, even if S raises the probability of X from (say) .4 to .45[9], the display carries information about X obtaining (though in this case a small amount). It is also a feature of this understanding of correlational information that it does not require S to make it more than 50 per cent likely that X obtains. The presence of S just needs to raise the probability of X. Even if the conditional probability of the state X given S is less than 0.5 the signal will still carry information about X obtaining. It will do so if the unconditional probability of X absent S is lower than the conditional probability of X given S.

Generally speaking, correlational accounts of content carry with them the least amount of theoretical baggage. This is because, unlike subsequent theories to be considered, correlational accounts do not depend on philosophically contentious notions of biological function. This austerity has its price, however. The price is the notorious problem(s) of content indeterminacy.

- Problem: indeterminacy -

Purely correlational accounts of content face indeterminacy problems. A leopard alarm call transmits content about the presence of a leopard because there is a correlation between leopard alarms and actual leopards. On the one hand, there is the probability of a leopard being present independent of an alarm call. On the other hand, there is the probability of a leopard being present given a leopard alarm call is given. The latter probability is higher than the former. However, correlation is cheap. Thus, we ought to worry that the content of the 'leopard' alarm call (individuated in

[8] It must be noted that Shea (2007) employs his notion of correlational information within a teleosemantic view of content.
[9] Although it is likely to carry significantly more information.

virtue of its physical structure) will be legion if content is assigned via a correlational theory.

For example, a leopard alarm call correlates with the sender being in a particular state of belief, just as well (if not better) than it correlates with the presence of an actual leopard[10]. So why isn't the content of the 'leopard call' the following: *sender believes there is a leopard here now*? Similarly, leopard alarm calls correlate with leopard-caused light patterns in the environment just as much as they correlate with actual leopards. Whenever there is an actual leopard in the environment there are always leopard-caused light patterns in the environment. This is the 'depth problem' facing informational semantics. The depth problem is one form of indeterminacy facing a purely correlational theory of content. It exists due to the basic fact that any signal that raises the probability of the presence of some middle-sized distal object in the environment like a leopard will also raise the probability of the proximal states used to detect and represent a leopard.

Even more worrying is the difficulty of ruling out wildly disjunctive contents. We might respond to the depth problem following Dretske. Dretske (1981) anticipated the depth problem and proposed a first-pass solution. The solution involves a sender that conditions production of a sign S (i.e. a leopard alarm call) carrying content C (i.e. *leopard here now*) on multiple proximal cues: P1 (i.e. leopard-caused light patterns) and P2 (leopard-caused sound waves). As a result of signal production being conditional upon the receipt of more than one proximal cue, a receiver cannot tell from S whether either P1 or P2 obtains. However, the receiver can tell from S whether C obtains. S carries much more information about C than either P1 or P2 alone. As a result (Dretske argued), C is the content of S as opposed to either of P1 or P2.

There is a problem, however. It is true that one cannot reliably infer one of either *P1* or *P2* from S, but unfortunately one can infer the disjunction *P1 OR P2*. Take the disjunctive set of proximal cues that variously elicit sign production, and S will carry just as much information about this disjunction as it will about C. Take the disjunctive world-state *leopard-caused light patterns* OR *leopard-caused sound waves* OR *beliefs about a leopard being here now*. A leopard alarm call will carry just as much information about this disjunction (appropriate filled out) than it will about actual leopards. Indeed, the alarm call might carry more information about this disjunction than actual leopards, because of false alarms. Why is this problematic? Because of my desideratum. The disjunctive state presented in this paragraph is true if any single one of its disjuncts is true. For instance, it will be true if the sender believes a leopard is

[10] No doubt, leopard alarms are sometimes given mistakenly. Thus, the signals will raise the probability that the sender believes there is a leopard here now more than it will raise the probability that there is a leopard here now.

present (the last disjunct). But the fact that the sender beliefs a leopard is present does not, by itself, rationalise the receiver's response to the alarm call. Only the presence of a leopard does this.

Dretske's (1986) second-pass solution to the depth problem was to invoke learning. If the set of proximal cues is open-ended over ontogenetic time, with the sender being able to condition its production of S on novel cues that indicate C, then S will carry more information about C as opposed to any particular set of proximal disjunctions at time T. However, a rather obvious problem with this second-pass solution to the depth problem is that it excludes 'hard-wired' communication systems. Many between-organism communications systems are developmentally canalised, with sender and receiver strategies remaining relatively fixed over ontogenetic time. But we don't want to deny content to these systems just because of this. Biological practice attributes content to canalised firefly mating signals just as readily as to vervet alarm calling[11], and both plastic and hard-wired signals pose the same explanatory challenge: why do receivers respond? So, some other solution to the depth problem is needed for the present project.

There is a second form of indeterminacy facing purely correlational accounts of content. This is the 'breadth problem'. Leopard alarms given by vervets are no doubt occasionally produced in error: for instance, in the presence of rustling bushes when a leopard is not, in fact, present. Thus, 'leopard signals' correlate more reliably with the disjunction *leopard OR rusting bush* than they do with just leopards. Needless to say, this is not a good outcome for a theory of signal content. The signal would mean *leopard present* OR *no leopard present/a bush is rustling*[12].

The breadth-problem is particularly pressing when sender and receiver interests begin to diverge, or conflict entirely. In such cases, we might expect signals to be given 'deceptively' at least some of the time. Like Searcy & Nowicki (2005) and Skyrms (2010), I use a functional definition of deception: 'deceptive' signals are those that raise the probability of a state of the world that does not, in fact, happen to obtain; such that the sender benefits at the expense of the receiver. Usually, deception can exist so long as the proportion of dishonest signals is not too high. Take the notorious example of firefly mating signals. Females of one species, *Photinus*, produce flashes of light to attract males seeking to mate. However, females of a closely related genus, *Photuris*, mimic the flash of *Photinus* females. They lure males of the latter genus in and eat them. Thus, the 'mating call' raises the probability that there is a female of the genus *Photinus* looking for a mate *or* that there is a female of the genus *Photuris*

[11] Which is developmentally canalised to a significant degree, at least when it comes to the sender's side of the interaction (Seyfarth & Cheney 2010).

[12] Even worse, if a vervet call raises the probability that a leopard is present, it equally raises the probability that a leopard is present *or* a typewriter is present (or anything else). The probability of *p* is never higher than the probability of *p* or *q*, for any *q* at all.

looking for a meal. Of course, disjunctive contents are not to be ruled out a priori. If the receiver engaged in some checking or probing behaviour before coming in to land, then it would right to say that the disjunctive content explained the response. However, without this, the latter content explains why the receiver responds as it typically does, as opposed to content about *Photuris* imposters. Moreover, only if content is non-disjunctive and relates solely to *Photinus* females willing to mate can the signal be used dishonestly or deceptively when given by *Photuris* females looking for a meal.

One conceivable response is this. It is to simply stipulate that the content of signals must be: a) non-disjunctive; and moreover b) that (non-disjunctive) world-state, the probability of which has been raised most by the signal. In our working example, this state will be *Photinus looking for a mate* and not *Photuris imposter looking for a meal*. This response might be acceptable under highly idealised conditions, such as for modelling purposes. Outside of such contexts, though, the response seems rather ad-hoc. I conclude, then, that correlational information faces a major problem as a candidate for animal signals: the problem of non-arbitrarily identifying a privileged world-state from among the many world-states a signal correlates with. This problem occurs on both the 'horizontal' dimension (depth) and the 'vertical' dimension (breadth).

These sources of indeterminacy are problematic because of our desideratum for control attribution. Content should make sense of why receivers respond to signals. Take the horizontal dimension. The fact that, say, a leopard alarm correlates with a leopard being present *or* the fact that the sender believes there is a leopard present, doesn't rationalise tree-climbing in response to the signal. This is because the content *leopard here now OR beliefs about a leopard here now* is true when there is just a belief about a leopard but no leopard. But in those instances where there was leopard belief in the absence of a leopard, climbing a tree in response to the signal did not contribute to this response being maintained. Similarly, the fact that a firefly mating call correlates with a female looking to mate *or* the sender wanting to eat the receiver doesn't rationalise the male's typical response. This is because the content *female looking to mate OR sender wanting a meal* is true when there is merely a sender wanting a meal. But when the sender wanted to eat the male as opposed to mate with him, the male approaching the sender was not a response that was reinforced by selection. It should also be noted that indeterminacy problems have also worried some biologists. Carazo & Font (2010, pp. 661-662), mirroring Rendall et al. (2009), state that a significant problem faces informational definitions of animal communication based simply on 'uncertainty reduction'. The problem "lies in specifying, of all the information potentially available to the receptor in a communicative interaction, the subset that should be considered relevant to the communicative context".

2.2 Input-orientated (causal-informational) teleosemantics

In response to these problems with causal-informational theories of content, a popular solution has been to go 'teleosemantic' and appeal to the notion of selected function. One variant of teleosemantics, to be addressed now, is more of an augmentation or modification of causal-informational semantics than an essentially novel kind of theory. It retains the notion fundamental to causal-informational semantics that content depends on correlation, and simply adds teleology to the mix to whittle down the promiscuity of correlation on both horizontal and vertical dimensions. On this 'input-orientated' teleosemantics, the content of a representation is what the representation correlates with when the representation producer is functioning correctly. For example, while a vervet alarm might correlate with a leopard *or* with beliefs (in the sender) about a leopard, it is the selected function of the signal to correlate only with a leopard being present (Dretske 1986, Neander 2013).

'Output-orientated' teleosemanticsts, to be considered in Section 2.3, have objected to the idea that the selected function of a representation can be to correlate. They argue that the selected effect of a trait must be forward-looking, or something that occurs causally-downstream of the trait's activation and in response to the trait's activation. But correlation is an upstream relation: it concerns the relationship between a trait and what caused the trait's activation. Because of this, it can't be the selected function of a trait to correlate with some world-state. This is because selected functions of traits are effects that traits have control over, and a trait has no control over what activated it: only over what *it* activates. In order to defend the idea that upstream correlation can be the selected function of a representation, Neander (2013, 2017) has arguing convincingly for the existence of 'response functions': systems that have a response function were selected to bring about some effect *in response* to particular upstream conditions. It is surely an important historical fact about pain receptors, for example, that they fired *in response* to biologically damaging stimuli, as opposed to biologically beneficial stimuli. Similarly, when it comes to animal communication. It is surely an important historical fact about communicative interactions that senders produced signals *in response* to appropriate world-states. In this light, it is more accurate to ascribe response functions to the systems that produce representations, as opposed to those representations themselves. When it is said that a leopard alarm call indicates the presence of a leopard and not the presence of a leopard *or* some other irrelevant world-state, because that is what the representation is supposed to do, what this really means is that the sender is supposed to token leopard alarm calls in the presence of a leopard. With this in mind, we will examine how input-orientated teleosemantics can solve indeterminacy problems.

First, the depth problem. Recall the general form of the problem facing a purely correlational theory of content. A vervet call pre-theoretically labelled as a 'snake alarm' raises the probability that a snake is here now. Thus, the content of the call is *snake here now*. However, the vocalisation also raises the probability of many world-states more proximal to the sender than the presence of actual snakes. It raises the probably that snake-caused patterns of light are present, and it raises the probability that beliefs about snakes being present are here now. Thus, the content of the vocalisation is unfortunately *snake OR snake-induced retinal firings OR beliefs about snakes*. Neander (2013, 2017) responds that there is a principled way to restrict content to the distal world-state. Adapting her argument to vervet signalling, we appeal to counterfactual asymmetries. The sender was selected to produce the snake alarm in response to snakes. It was selected to do this *via* producing the alarm in response to snake-caused light patterns. However, the sender was not selected to produce the alarm in response to snake-caused light patterns *via* producing it in response to snakes. As a result, we restrict content to the more distal world-state of *snake*.

What about the breadth problem? The problem, to recall, is that pre-theoretically labelled 'snake alarms' are sometimes mistakenly given in response to distal objects that aren't snakes, such as snake-like sticks. Thus, 'snake alarms' carry the content *snake OR snake-like stick*. Additionally, vervets have been reported to occasionally give an alarm call in the presence of an individual attempting to migrate between groups, to stop the migration (Dennett 1998). If cases of deception like this occasionally occur, tokens of the signal-type would mean *snake OR snake-like stick OR attempted migration*. However, if we focus on the selected function of the sender, the problem just raised for mistaken signals, but not deceptive signals, can be bypassed. The idea is as follows: it is the selected function of senders, qua response systems, to token 'snake alarms' in the presence of snakes as opposed to snake-like sticks. Thus, the content of a 'snake alarm' is *snake* and not *snake-like stick*. However, causal-informational teleosemantics doesn't do as well on the breadth problem when it comes to deception.

- Problem 1: non-cooperative communication -

Consider alarm calls produced deceptively. Presumably, it is an adaptation of a sender (even if over ontogenetic time) to produce an alarm sometimes in the presence of an undesired migration attempt. If this is right, then there is a problem for causal-informational teleosemantics. On the latter, so-called 'snake alarms' will carry content about both snakes *and* intragroup migrations. This is a problem given the desideratum for content-attribution suggested in Section 1. Content must explain why receivers respond to signals. But the disjunctive content *snake OR intragroup migration* doesn't explain why receivers typically respond to snake alarms as they do. This is because the disjunctive content is true when there is just a migration occurring but no snake. But in such cases the typical response was not rewarded. In contrast, the content *snake* explains why receivers typically respond to snake alarms as they do.

Another problem is that, given a causal-informational teleosemantics, it cannot be said that so-called 'snake alarms' are ever tokened deceptively. This is because the so-called 'snake alarm' isn't just a snake alarm, under the theory. It's also a migration alarm. This is because the tendency to produce putative 'snake alarms' in response to snakes or migration attempts is an adaptation of the sender. Thus, the signal can't be tokened deceptively when a migration is taking place.

- Problem 2: adaptation and indication -

The fundamental idea behind input-orientated teleosemantics is that the content of a signal-type S is determined by focusing on what producers of S have the selected function of indicating. I want to focus on some important commentary on this idea given by Godfrey-Smith. Godfrey-Smith (1992) has argued that indication is a problematic answer to the question *what is the selected function of senders, qua producers of S?* He argues that S can represent X even if the statistical relationship between S and X is quite weak. X just needs to be a world-state that is important in the selection of S as a cause of the response S typically elicits in a receiver[13].

Consider detection tasks where the costs of false negatives far outweigh the costs of false positives. Take an acoustic vocalisation representing the presence of a predator. Was this acoustic pattern recruited by the receiver as a signal of a predator because a*) whenever it was produced, a predator was lurking*, or alternatively because b) *whenever there was a predator lurking, the signal was produced*? These two possibilities are quite different. If the former is what explains why the signal was recruited by the receiver as an elicitor of response R, then an explanation of why the signal was recruited may well involve the idea that it indicates a predator. However, if the latter is what accounts for the sign's recruitment, then an explanation of why the sign was recruited need not involve the sign correlating with a predator. For *this*, latter, task can be carried out more effectively by the sender producing signals when there is no lurking predator around. That is, it can be carried out more effectively by the sender producing many false positives. By creating many false positives, a sender can actually better ensure that there is a sign *whenever there is a predator nearby*.

A proponent of indication as the selected function of a sender might respond that, despite the positive statistical relation between S and X being weak in this case, it

[13] Readers in favour of Millikan's output-based theory of content may begin to be feeling more at home here.

exists nonetheless. They could point out that S still raises the probability of X, just by a small amount. Perhaps 80% of S tokens (the predator alarm) are given in the absence of X (the appropriate predator), such that the chance of X, given S, is a low 20%. Still, if the prior probability of W obtaining is even lower at 5%, the fact that 20% of S tokens are given in the presence of X means that S raises the probability of X: by 15% to be exact. So S still correlates with X, where this correlation explains why S was recruited by receivers as a stand-in for X. However, we can modify the case such that it becomes problematic to say that the function of a sender (or a signal) is to convey information about X.

Suppose that, as above, S raises the probability of some relevant world-state X (i.e. that a predator is present) by 15%: e.g. from 0.05 to 0.2. However, assume now that S also raises the probability of another world-state X^* (say, that an individual of a benign species which looks rather like X individuals is present). Assume, further, that S raises the probability of X^* (a benign look-alike) by more than it raises the probability of X. That is, individuals of the look-alike species cause tokens of S more often than individuals of the predator species do. Now, suppose that the cost of initiating predator-avoidance behaviour in the presence of a look-alike is quite minor. Suppose also that S is the best indicator of X (actual predators) available: no more fine-grained sensory discrimination on the part of the sender is possible. In this case, S will be recruited as an indicator of X despite carrying more information about X^*. The problem with this is that it now seems strange to say that S was recruited because it correlates with X. What seems more appropriate to say is that S was recruited because it elicits behaviour in a receiver that is beneficial given the statistical relation between S and X *but also given a host of other factors*: such as the relevant payoffs and the relevant constraints.

It arguably gets worse for the notion that senders (or signals) are selected for correlating with their contents. The above case is one in which S is the closest thing possible to a genuine 'indicator' of W. Despite S carrying more information about X^* than X, the ideal is still that S *would*, if it could, carry information about X more reliably than X^*. It's just the case that, for whatever reason, constraints got in the way of this ideal. If, contrary to fact, the sender was able to discriminate between Xs and X^*s, then a signal would eventuate that correlates more strongly with W than with X^*.

However, this might not be true. That is, receivers might sometimes *not even want* a signal to correlate more strongly with X than with some non-X even if this were possible. This would be the case if the cost of error is huge. If, whenever a sender fails to notice the presence of an actual predator, there is a large chance that the receiver will be killed and if the cost of climbing a tree in response to a benign look-alike is quite negligible, then it could well be more adaptive (from the point of view of the receiver) for S to correlate more reliably with X*s than with Xs[14]. In this case the selected function of the sender (or signal) would seem to be to correlate with X*, not X!

The ultimate point is that it isn't reliable correlation that is being selected for in this kind of case. The statistical relation between signal and world is only part of a complex evolutionary explanation of a sender's production strategies. Equally relevant are constraints, and, crucially, costs and benefits. As already noted, a proponent of input-orientated teleosemantics could concede that the signal does not correlate most reliably with the relevant world-state X but maintain that the correlation with X is nevertheless relevant: 'sure (they might reply) S carries more information about a world-state that isn't its content, but it still raises the probability of its content, even if less than other states!'.

I do not intend to settle this issue once and for all here. However, it does seem strange to me that S can have the function of indicating X no matter how strong or weak the positive statistical relation between the two. One consequence of this theory will be that *more information transmission isn't necessary better*. This is slightly strange. If the function of the opposable human thumb is to enable the grasping of objects, it would seem that any variation in the thumb that enabled an individual to better grasp objects would be favoured, given relevant developmental trade-offs. However, given relevant trade-offs, a signal that correlates more strongly with its content is not always more favourable than a less informative signal. Although perhaps not definitive, this consequence of input-orientated teleosemantics at least calls into question the idea that it is the function of senders to correlate signals with world-states.

This seems like a good place to stop and take stock. By augmenting purely correlational theories of content with the idea that it is the selected function of senders to correlate, various problems can plausibly be solved. First, Neadner's (2013) functional asymmetry considerations plausibly solve the depth problem. Second, looking at what state the sender is *designed* to raise the probability of by signalling might solve the breadth problem. We saw that while input-orientated teleosemantics might be able to handle signals sent accidently in the presence of the 'wrong' world-state, it has trouble assigning appropriate contents when signals are sent *deceptively* by the sender: i.e. when the non-'veridical' tokening of a signal is an adaptation for the sender but detrimental to the receiver. A more fundamental objection to input-orientated teleosemantics was then considered. This objection questioned whether it gets the functional considerations of communication right to say that senders were selected for correlating signals with upstream world-states.

---

[14] At the very least, it would be more adaptive for S to correlate more reliably with the disjunctive world-state *W or W** than with W alone.

2.3 Output-orientated teleosemantics: 'Biosemantics'

Now we will explore output-orientated teleosemantic theories of content, specifically as they apply to animal communication. All have in common the idea that indeterminacy problems can be solved by focusing on the selected function of the receiver. Whereas input-orientated teleosemantics stresses the function of the sender (which, in the theory, is to correlate signals with world-states), biosemantics stresses the function of the receiver. What determines content for biosemantics is "what the sign needs to correspond to if the consumer is to perform its tasks in its normal way" (Millikan 2004, p. 80). The content of a signal is to be identified with the world-state that rationalises (in a fitness sense) the behaviour elicited in the receiver.

The first step in determining content is to discover what the receiver was selected for doing in response to some type of signal. The second step is to determine the 'normal condition for the proper performance' of the receiver's response. Take the 'mating call' of the *Photinus* female firefly, also sent deceptively by female *Photuris* fireflies. The receiver is a male *Photinus*. The function of the male's response to the signal is to bring the male closer towards the female. Now, this response only makes evolutionary sense (from the perspective of a male *Photinus*) when the sender is a *Photinus* female. When the male caller is a predatory *Photuris* female, the behaviour of a male is not rewarded by selection. Approaching makes no evolutionary sense. Thus, the normal condition for the proper performance of the male's response is that the sender is a *Photinus* female looking to mate, and the signal is genuinely a mating call, as opposed to something less determinate.

The virtues of this kind of theory of content are apparent when contrasted with causal-informational semantics. First, the depth problem. Even though leopard alarms correlate with, say, leopard-induced retinal firings as well as leopards being present, the content of leopard alarms refers to leopards being present and not leopard-induced retinal firings. This is because the former, but not the latter world-state, makes evolutionary sense of the receiver's response to the signal (i.e. tree-climbing). Next, one variant of the breadth problem. Imagine, plausibly, that vervets sometimes produce 'leopard' alarm calls mistakenly, perhaps in response to a rustling bush. As a result, the 'leopard' alarm will correlate more reliably with *leopard present OR no leopard present/rustling bush* than with *leopard present*. However, the behaviour elicited in a receiver by the signal is tree-climbing. And tree-climbing behaviour only makes evolutionary sense in the presence of a leopard, as opposed to rustling bushes in the absence of a leopard. Thus, content is restricted to *leopard here now*.

Biosemantics also does better than input-oriented teleosemantics in handling deception. Even though predatory *Photuris* females sometimes give the *Photinus* mating signal, such that the signal correlates with *Photinus female looking to mate OR Photuris* *female looking for a meal*, this fact is irrelevant in fixing the content of the call. Input-oriented teleosemantics struggled because, not only does the deceptive signal correlate with a disjunctive world-state, it does this, also, as a matter of proper function. *Photuris* females have been adapted to produce the *Photinus* mating signal[15]. However, biosemantics is not troubled because only the presence of a *Photinus* female looking to mate makes evolutionary sense of the receiver's response to the signal (i.e. approaching the sender). Similarly, biosemantics can accommodate vervet alarm calls given during an attempted migration. In this case, the success condition of the predator-avoidance behaviour of receivers is that there is a snake here now, and *not* that a migration is occurring. Thus the content is *snake* as opposed to *snake here now OR attempted migration*. Causal-information teleosemantics ran into trouble because, again, not only do alarm calls given deceptively correlate with the disjunctive world-state, alarm calls do this as a matter of proper function.

- Problem 1: evolutionary success conditions are indeterminate -

Biosemantics is thus able to solve the indeterminacy problems facing a purely correlational theory. It is even able to improve upon input-orientated teleosemantics when interests of senders and receivers conflict. However, biosemantics faces problems of its own (See Neander (2012) for an overview). One issue, which I will not address here, originates with Fodor and is motivated by a blanket scepticism about the ability of natural selection to select *for* particular traits and not other co-extensive traits. It has been responded to in detail elsewhere and so I will not address it here[16].

The first indeterminacy problem for biosemantics to be addressed here again originates with Fodor (1990) but is less extreme. It is the problem of arbitrarily-described evolutionary success conditions. Natural selection cares about reproductive success, but not the particular description we use to explain this success. When it comes to communication, whether a signal gets the receiver to produce its typical response in the appropriate circumstance is visible to selection, but the description we choose for the world-state that rationalises this response is largely arbitrary. Does the 'leopard alarm' mean *leopard*? *Predator*? How about *survival and reproduction hindrance*? While some of these aren't intuitive candidates for the content of this signal-kind, each is perfectly plausible given the theoretical machinery of biosemantics: the presence of any one of these world-states rationalises the response of the receiver to leopard alarms.

[15] Of course, I am assuming that the *Photuris* signal is the same kind of signal as the *Photinus* signal. I will give reasons for this assumptions soon.
[16] Neander (2005), Rosenberg (2013), and Sterelny (1990) have each responded to Fodor's anti-Darwinian scepticism about naturalising content.

The second is that of omnipresent evolutionary success conditions. Many things in addition to the presence of a leopard were required on past occasions when leopard alarms functioned properly: oxygen in the atmosphere; gravity operating normally (and so on). Because biosemantics identifies content with the evolutionary success condition of the behaviour elicited by the signal, content will be *leopard here now in the presence of sufficient oxygen in the atmosphere and gravity operating normally* (and so on). The content for the 'leopard alarm' will be a potentially limitless list of success conditions and so completely unwieldy.

The third is the problem of overly-specific evolutionary success conditions. All sorts of things could have prevented vervets from gaining a fitness benefit from the leopard alarm: e.g. limb injury preventing a vervet from climbing a tree successfully in response to the signal. Again, because content is identified with the evolutionary success conditions of the response elicited by the signal, the content of the 'leopard alarm' will be *leopard here now and the receiver isn't badly injured*. Likewise, imagine a crucial branch breaking while a vervet is climbing a tree in response to the signal. In such a case tree-climbing behaviour would not have earned its keep. Thus, the content of the 'leopard alarm' must be *leopard here now and the receiver is able to climb and the receiver will not choose a weak branch for a perch*. The list of specific success conditions is potentially endless, meaning that, again, the content of the 'leopard alarm' will be extraordinarily detailed.

Shea (2007) offers a solution the first problem: that of arbitrarily-described evolutionary success conditions. His solution is designed to allow biosemantics to rule out all of the candidate contents listed above (*leopard; predator; survival and reproductive hindrance)* apart from *leopard*. His proposal is as follows:

> The evolutionary success condition relevant to a particular representation should be specific to the behaviour prompted by that representation, as compared with behaviour prompted by other representations that mediate between the same producer and consumer system (Shea 2007, p. 408).

While there might be a very general success condition relevant to the responses elicited by all signal-types, there is a more specific success condition relevant to the response elicited by each particular signal-type. This latter, more specific, success condition is the content of the signal. While *survival and reproduction hindrance* is a world-state that explains why all the predator-avoidance behaviours of vervets (in response to alarm calls) are advantageous, it is not a world-state rationalising tree-climbing (in response to leopard alarms) as opposed to bush-hiding (in response to eagle alarms). Similarly for *predator*: this is not a world-state rationalising tree-climbing as opposed to bush-hiding. Out of the normal candidates, only *leopard* does this. Of course, Shea's proposal depends on there being more than one type of signal. What about single signal systems, like firefly mating flashes? Because this communication system is just signal/no-signal, each signal would naturally mean any of the following: *female Photinus looking to mate; reproduction-enhancer looking to mate; something that won't kill you looking to mate*, etc.

In response to this I am inclined to bite the bullet. When we are considering non-linguistic signalling systems like firefly flashes or vervet alarms, the expectation that a theory of content deliver contents that are as fine-grained as linguistic utterances is, I think, unrealistic. In addition, the candidate contents of the *Photinus* mating signal all genuinely explain why receivers respond. Thus, while it doesn't deliver contents as fine-grained as linguistic signals, biosemantics is still a genuine improvement over causal-informational semantics where many candidate contents do not make sense of why receivers respond. It is one thing for a signal's content to be indeterminate between *female Photinus looking to mate* or *something that won't kill you looking to mate*. It is another thing for content to be indeterminate between *female Photinus looking to mate* or *female Photuris looking for a meal*. The latter content goes awry of our desideratum for content attribution; the former does not.

What about the second and third problems: omnipresent and overly-specific evolutionary success conditions? Both seem to be genuinely problematic for biosemantics. As I will argue in Section 3, though, a variant of biosemantics known as 'infotel semantics' (Shea 2007) has the resources to neutralise both indeterminacy problems without biting any bullets. Before addressing this, though, I address another problem biosemantics faces in the context of animal signals. This is the problem of non-cooperative communication again. I explain the problem as it applies to biosemantics before arguing that a recently-proposed solution for biosemantics fails (Artiga 2014). I then argue that tweaking biosemantic theory along the lines suggested by Stegmann (2005a, 2009) handles the challenge from non-cooperative communication[17].

- Problem 2: non-cooperative communication (again) -

As it did for input-orientated teleosemantics, deception again poses a problem for teleological theories of content. Only this time the problem confronts biosemantics. Deception is problematic for biosemantics because biosemantics, at least as traditionally conceived, is a cooperative theory of representation: the proper function of the sender (or receiver) is part of a normal explanation for why the receiver (or sender's) evolved act was adaptive. This is a fundamental component of the theoretical machinery of biosemantics (Millikan 2004).

[17] And because infotel-semantics is itself a variant of biosemantics, Stegmann's modification to biosemantics can be carried through to infotel-semantics, as will become clear soon.

To see why biosemantic theory breaks down under conditions of non-cooperation, take the example of the hungry *Photuris* female firefly preying on a hopeful *Photinus* male looking for a mate. On Millikan's theory, the content of a signal is determined by (a) what the sender is designed to map via its signal production strategies, where (b) this mapping enters into a normal explanation of why the receiver's stereotypical response is adaptive. But in the case of the predatory mimic, (a) will not gel with (b) appropriately. More specifically, the world-state that the sender evolved to map via producing the signal is not the world-state that explains why the receiver's response evolved. For the predatory mimic maps the fact that there is a hungry predator looking for a meal. But the fact that the signal was produced by a hungry predator looking to eat the receiver definitively does not explain why the stereotypical response of the male receiver (namely flying closer to the sender) evolved! The male's response evolved when the sender was a female non-mimic willing to mate, not when the sender was a predatory mimic of a closely related species looking to eat him. So there is a problem for Millikan's biosemantic theory of content. Non-cooperative communication throws a spanner into the theoretical works, such that the machinery breaks down.

In response to this problem, Artiga (2014) comes to the aid of Millikan's biosemantics. He argues that biosemantics can successfully assign content when interests conflict, because the mimic's signal belongs to the same biological kind as the honest signal. Recall that aggressively mimicry is prima facie problematic for biosemantics because the latter requires the proper function of the sender's act to be a normal condition for the proper performance of the receiver's act. But in cases of aggressive mimicry, the proper function of the sender's act (i.e. to secure a meal) is not a normal condition for the proper performance of the receiver's act (i.e. being attracted by the signal). According to Artiga, however, we can class the two signals as being of the same kind, thus removing the conflict between proper function of the mimic sender's act, on the one hand, and the historical success conditions of the receiver's response, on the other.

The idea is there is a relationship of counterfactual dependency between mimic signal and honest signal, such that the form of the mimic signal tracks (over phylogenetic time, and no doubt with some lag) variation in the honest signal. The mimic signal and the honest signal "belong to the same kind in virtue of the fact that they tend to have many properties in common due to an underlying robust causal mechanism" (Artiga 2014, p. 370). There is a 'homeostatic clustering' mechanism at work, which ensures that the mimic signal is shaped by the cooperative relationship between the honest sender and receiver just as much as the honest signal is. Thus, tokens of the mimic signal belong to the same kind as tokens of the honest signal. I think this proposal fails. The mimic signal cannot belong to the same kind as the honest signal due to counterfactual dependencies. This is because the counterfactual dependencies only go one way. Change in the honest signal tends to cause change in the mimic signal, true. However, this doesn't hold in reverse. Change in the mimic signal does not cause change in the honest signal. Moreover, the honest signal is under selection to diverge from the mimic signal.

Helpfully, Stegmann (2005a, 2009) has modified biosemantics to accommodate non-cooperative communication. His proposal departs from Millikan's essentially cooperative biosemantics in being exclusively receiver-focused. Both representational status and content are grounded independently of the sender. Indeed, a sender need not even exist for something to be a representation with content. Some biological entity is a representation, for Stegmann, if it is used by a consumer to respond adaptively to the world. And the content of the representation is the *particular* state of the world that explains why the consumer's response to the biological entity is (normally) adaptive. On Stegmann's proposal, assigning content to the mimic's behaviour in our example would go as follows. The mimic's flashing behaviour is a representation because it is used by *Photinus* males looking to mate with a *Photinus* female. The content of the representation is the particular state of the world explaining why *Photinus* males respond to the flashing display by becoming sexually receptive and flying closer. Clearly, this state is that there is a *Photinus* female signalling who is looking to mate with a male of her own species. So far so good. However, one might worry about how we can group the mimic's deceptive flash alongside a non-mimic's flash. This needs to be possible so that the receiver's response to the mimic's flashing display inherits the historical success conditions of receiver's responses to the honest display. All theories of content inherit this problem. However, one possibility is to group flashes in terms of their selected function and causal properties relevant to the fulfilment of this selected function: all traits with the selected function of attracting a male *Photinus* looking to mate, and that have properties *P1,…, Pn* causally relevant to attracting a *Photinus* male looking to mate, inherit the historical success conditions of a male *Photinus*' typical response (Stegmann 2009). If this proposal works, then a wholly consumer-based teleosemantic theory of signal content can successfully assign content to both cooperative signals such as vervet alarm calls and also non-cooperative signals like aggressive mimicry among fireflies.

Before moving on, Artiga (2014) has argued against Stegmann's proposal in clearing the way for his own proposal, which I argued above fails. These objections need to be neutralised before running with Stegmann's biosemantics. Artiga's concern is that focusing exclusively on the receiver solves one problem at the expensive of opening up at least 3 others. Firstly, Stegmann's view is allegedly problematic "not only because it changes the framework for thinking about signals, but also because it compels us to reconsider the use of abstract models of signaling in analyzing the phenomenon of representation[18]" (Artiga 2014, p. 365). This objection fails because it groups together two issues which are conceptually separate: classifying certain traits as signals, on the one hand, and assigning specific contents to signals, on the other. Stegmann's theory of content has it that cues, for example, will carry content. But this doesn't make a cue a signal. Stegmann himself (2009, p. 872) writes:

> Signals are not distinguished from cues by the kind of information they carry (i.e., semantic vs natural information). Rather, as current biological science suggests, signals are evolved for the purpose of representing and carrying information, whereas cues are not[19].

Artiga's second, closely related, charge is that representation will be promiscuous on Stegmann's account. According to the latter, an item is a representation if it "is consumed by a mechanism whose normal condition to perform its function is that items of this type correspond to something other than themselves according to a mapping" (Artiga 2014, p. 365). As a result, smoke will qualify as representations. This is because smoke is often used by an organism to stand in for the presence of fire, according to a mapping rule. Indeed all the sensory information we use in our day-to-day lives will be representational, for the same reason. In response, though, we need not take Stegmann's theory of content to determine what is and what is not a representation, let alone a signal. The representational status question is a thorny one (Ramsey 2007, Burge 2010). However, it is not the same question as what contents should be ascribed to a representation (let alone a signal), after a structure in question has been classified as a representation (or a signal). I propose using Stegmann's account of content to assign natural meaning to a trait that is independently considered a signal (as opposed to, say, a cue). Then, the content of a signal is the success condition of the elicited response. But this need not entail that a signal (or representation) is anything and everything eliciting an evolved (or learnt) response from an organism. My proposal is that we use Stegmann's theory of content for animal signals, but only after we have identified some trait as a signal. If something doesn't count as a signal, then we don't apply Stegmann's theory of content to it.

Artiga's (2014, p. 366) third charge against Stegmann is that "any representation (token) has always as many contents as potential consumers". For instance, a leopard call in vervetese will be about the presence of leopards for vervets, but it will also be about the presence of *vervets* for leopards. This is because the success condition of the

[18] Such as the Lewis-Skyrms model.

[19] If one is not happy with the definition that signals (but not cues) evolved for the purpose of representing or carrying information, then one can modify the definition and retain the necessary distinction between cues and signals. Signals, but not cues, evolved for the purpose of influencing another organism. Of course, this definition will be far too liberal to differentiate signals from many other non-signals that aren't cues, such as acts of coercion. But, as Chapter 3 will show, no currently-available definition (whether it employs information, or merely influence) is unproblematic.

leopard's response (hunting the sender) is something like the fact that a tasty vervet produced the call. Put simply, there is no principled way of discounting eavesdroppers as receivers of the signal and, as a result, signals will have as many contents as consumers. But this runs afoul of our desideratum, which is that content should help make sense of why the receiver, picked out by contextual-pragmatic considerations, responds, and hence why signalling has not collapsed. The fact that a leopard responds to the signal, and responds because (for the leopard) the biosemantic content of the signal is something like *vervet here now*, doesn't help us understand why vervets respond.

But why should this be an objection that derails Stegmann's proposal? In one sense, leopards *are* receivers if they recognise and respond to vervet calls, and their response will be explained by the fact that, when leopards benefitted from responding to vervet calls, vervet calls mapped to the presence of vervets: in other words, the content of the signal, for leopards, will help us understand why leopards respond to the signal (in this case, the content *vervet here now* won't help us understand why senders produce the signal, though). Alternatively, one might simply treat the vervet call as a cue, from the point of view of a leopard. As already argued, it need not be the case that everything eliciting an evolved/learnt response from an organism counts as signal, even though our theory of content tells us to look simply to the success condition of the evolved/learnt response. On this proposal, a vervet call is a signal from the perspective of another vervet but a cue from the perspective of a leopard. It is a signal for vervets because of co-adaptation. It isn't a signal for leopards, because of a lack of co-adaptation[20].

- *Stocktake* -

A traditional (sender-receiver) biosemantic theory of content for animal signals faces a significant problem when the interests of senders and receivers conflict. An integral part of the biosemantic machinery is that the proper performance of the sender's function is a normal condition for the proper performance of the receiver's function. But when interests conflict, this won't be the case. In response, I defended Stegmann's proposal to focus exclusively on the receiver when it comes to assigning content. However, a biosemantic theory of content still faces indeterminacy problems, as outlined earlier: the problems of omnipresent and overly-specific evolutionary success conditions.

Furthermore, biosemantics faces *yet another* issue that needs to be addressed. This issue is serious given my desideratum for content attribution presented in Section

[20] Chapter 3 will address the thorny issue of defining communication.

1. The issue is that biosemantics arguably cannot adequately explain the success of responding to an accurate representation. This charge has been used to argue that biosemantics requires representations to correlate with their contents[21]. Only by correlating with their contents can accurate representations explain successful responses. I begin the next section by explaining the objection and relating it to between-organism signals, specifically. I will argue that the solution to this *problem of circularity*, as well as to the two outstanding indeterminacy problems for biosemantics, can be solved only by moving to a different theory of content: 'infotel-semantics'. Infotel-semantics draws heavily from output-orientated teleosemantics (i.e. biosemantics) but it (re)incorporates the idea that signals must correlate with their contents. By adding correlation into the biosemantic mix, infotel-semantics is able to overcome both the problem of circularity and the remaining problems of indeterminacy facing biosemantics.

3. Biosemantics & Explanatory Circularity: Resurrecting Correlation?

3.1 THE CIRCULARITY OBJECTION

> The only value of a signal to a receiver is as a source of information, and without this value the receivers will not respond to the signal and the signallers will not product it. (Searcy & Nowicki 2005, p. 207).

As argued in Section 1, the main methodological justification for attributing content to signals is to makes sense of why receivers respond to signals (and hence also why senders produce them). In this section, I address an allegation that has been levelled against biosemantics. The allegation is potentially problematic in light of what was argued in Section 1. The allegation is that if content is assigned via a biosemantic theory, then a particular kind of explanation of behaviour will be circular.

First, we must get clear on the kind of explanation the objection has in mind. The relevant kind of explanation is one that makes sense of 1) behavioural success by 2) invoking the receipt of (accurate) information from a reliable signal. For instance, take a situation in which we explain why Joe succeeded in not being hit in the head with a ball because his friend yelled 'Ball!', causing Joe to duck suddenly. On the one hand we have an explanandum: the success of Joe in avoiding the ball (by ducking). On the other hand we have an explanans: the fact that Joe's friend yelled 'Ball!'. More precisely, the explanans is the fact that Joe's friend transmitted the accurate

[21] It is vital to note that an output-orientated teleosemantic theory (i.e. a biosemantic theory) which includes correlation bis different from an input-orientated teleosemantics. As will become clear, the former, despite possessing a role for correlational information, still ultimately assigns content based on the selected response of the receiver. By way of contrast, an input-orientated teleosemantics (i.e. Dretske 1986/Neander 2013) assigns content based on the selected function of the signal/producer.

information that there was, in fact, a ball headed towards Joe which caused Joe to duck. In short, the relevant kind of explanation is one in which acting on an accurate representation explains success. When a receiver is influenced by a signal in some way (say, by ducking), success typically obtains when: (1) the signal has content X (i.e. 'Ball!') and (2) X happens to obtain (i.e. when there is, in fact, a ball). The charge levelled against biosemantics is that we are unable to give explanations of the above kind. Why? Because biosemantics is a kind of success semantics (Godfrey-Smith 1996). Arguably, all forms of success semantics suffer from the problem of circularity:

> For correspondence[22] to have a real role in the production and explanation of success, it must be conceptually distinct from the fact of success. Success-linked theories threaten this independence. (Godfrey-Smith 1996, p. 192).

The alleged problem is that success-linked theories of content make a definitional connection between representing accurately (or 'corresponding'), on the one hand, and success, on the other. Teleosemantics is allegedly no exception:

> Teleosemantics too makes a definitional connection between representing truly and succeeding, and the fact that the definitional connection relates only to past episodes of behaviour does not make the explanation of the success of current behaviour any more substantial. (Shea 2007, p. 192).

For example, on a biosemantic account of content a leopard alarm carries information about leopards in virtue of a normal explanation of the past success of the receiver's response. Such an explanation will appeal to a 'mapping rule' which relates actual leopards to the time and place in which the signal is given. The nature of this mapping rule is determined by what made it the case that, when receivers historically responded in the typical way to leopard alarm calls[23], such a response was adaptive/successful. In this case, what made it the cases that the response was adaptive/successful was that there was a leopard present. As such, the content of a leopard alarm call is something like *leopard here now*. The alleged problem for biosemantics concerns the fact that the content of the leopard alarm call is determined by conditions of historical success. If content is *defined* in terms of what state of the world made the response to the signal adaptive for receivers in the past, content can't then be cited in an explanation *of* success, without the explanation being circular and hence empty. An adequate explanation of why a vervet successfully avoids being killed by a leopard in climbing a tree cannot cite the fact that the receiver derived the content *leopard here now* from an alarm call and that the content was correct.

How exactly is this objection to biosemantics potentially problematic for the present project? Well, if, as argued in Section 1, the explanatory role of attributing

---

[22] Godfrey-Smith uses the term 'correspondence' where I use the term 'accuracy'.

[23] I.e. by climbing a tree or remaining in a tree if already in one.

content to animal signals is to make sense of why receivers attend to signals; and if, on a biosemantic theory of content, success in responding to a signal can't be explained by invoking accuracy, then biosemantics will not be a good theory of content for animal signals. I will argue in the next sub-section that this objection doesn't affect the current project. However, I will show that a closely related objection does.

3.2 Assessing the Objection

It is important to note that the goal of explaining success by invoking a response caused by an accurate signal is subtly different from the desideratum argued for in Section 1. Section 1 argued that content must explain why receivers respond to signals. But this is different from explaining success. To illustrate this, assume that the present environment for a group of vervets has changed, very recently, from the one in which receivers learnt the significance of a leopard alarm call. For instance, if tree branches are now unable to support the weight of a vervet monkey for some reason, the fact that a leopard alarm call was produced accurately (i.e. in the presence of an actual leopard) will not explain successful behaviour. Nevertheless, the fact that the signal carries content about leopard-presence can still explain why a receiver happens to respond to the signal by climbing a tree. The requirement to explain success (Godfrey-Smith and Shea's desideratum) is a requirement to explain *present* success. However, the requirement to explain why receivers respond to signals (my desideratum) is a requirement to explain *past* success. Past success, and hence why receivers are wired-up, in the present, to respond to signals in the way they do, is something I've claimed a good theory of signal content must illuminate. On the other hand, present success[24] is something Godfrey-Smith (1996) and Shea (2007) claim a good theory of content must illuminate. These two desiderata are subtly different. But the difference matters. The difference means that the circularity objection isn't a problem for biosemantics, given my project. If the explanatory role of attributing content to animal signals is not, in fact, to explain present success, then it will not matter if present success cannot be explained in non-circular fashion.

However, might it be the case that the circularity objection applies equally to explaining past success as it does to explaining present success? If this is the case, then a biosemantic theory of signal content might still be problematic. It will be problematic if the historical success of typical responses to signals, over phylogenetic or ontogenetic time, cannot be explained by citing the receipt of accurate signals in those historical environments. Take the mating signal of female *Photinus* fireflies. What is the content of the signal, according to biosemantics? To answer this question, we

[24] More precisely: present success in virtue of deriving accurate content from a signal.

first look to the species-typical response of the target receiver. The target receiver is a male looking to mate with a female. The species-typical response of the male is to become sexually receptive to the signalling female. Next, we ask what world-state obtained when this species-typical response to the mating signal proved adaptive to male ancestors of our present male. The answer is that the sender of the signal was a female *Photinus* firefly looking to mate. The content of the signal, according to biosemantics, is thus something like *here is a Photinus female looking to mate*. Now the important question is this: can the receipt of accurate biosemantic content explain why the male ancestors of our present male experienced a fitness boost by responding with increased sexual receptivity? More precisely, can we do so a non-circular fashion?

One should be sceptical. This is because, according to biosemantics, content doesn't exist independently of the selected function of the receiver's response. Only by consulting a normal explanation for the success of the receiver's response can we assign content. In other words, only *after* selection has operated does content come into existence. As Lean (2014) has noted (in a paper on the methodology of applying content to certain within-organism causes as opposed to external signals), this gets things the wrong way around. What explains the content carried by a signal is not the adaptive response of the receiver. Instead, what explains the adaptive response of the receiver is the content carried by the signal. Content is the explanans, not the explanandum. If this is right (and I think it is), then a slightly different form of the circularity objection is problematic for biosemantics from the perspective of my project. In the next sub-section, I show how adding a clause relating to correlational information gets around the problem of circularity. If a signal correlates with its content, then this correlation can explain historical success. However, the selected function of the receiver is still highly relevant, in order to circumvent indeterminacy problems that would otherwise plague a purely correlational account of content. The best theory of content needs its snout in both the trough of correlational information and the trough of (the) selected function (of the receiver).

3.3 Accommodating the objection: Uniting correlational information and biosemantics

Enter 'infotel-semantics' (Shea 2007). Infotel-semantics is like biosemantics in a) stressing the selected function of the consumer and b) identifying content with the normal condition for the proper performance of this function. However, it brings correlational information back to the forefront of a theory of content. According to infotel-semantics, a signal of type S has content C if:

> Ss are intermediates in a system consisting of a sender and a receiver cooperating by means of a range of mediating signals (all specified non-intentionally), in which every signal in the range also satisfies (a) to (c);
>
> a) Ss carry the correlational information that condition C obtains
> b) An evolutionary explanation of the current existence of the signalling system adverts to Ss having carried information about C
> c) C is the evolutionary success condition, specific to Ss, of the behaviour of the receiver prompted by Ss. (Shea 2007, p. 419)[25]

Condition (a) stipulates that tokens of S must raise the probability of C in order to represent C. Condition (b) stipulates that S must have carried correlational information about C whenever the signal proved its worth over phylogenetic or ontogenetic time, where c) C is the world-state that 'rationalised' (in terms of fitness) the response elicited by S in the receiver.

What explains *why* receivers attend to signals is clause (b). The fact that signals historically correlated with the success condition of the receiver's response explains why the receiver's response has been maintained by selection. Moreover, because the selected function of the receiver is also appealed to in clause c, infotel-semantics can get around the problem of non-cooperation communication facing causal-informational teleosemantics. Recall that the problem for causal-informational teleosemantics was to assign the right contents when sender interests are opposed to receiver interests. Because content is assigned by looking at what world-state the sender was selected to produce the signal in response to, cases of deception are problematic. For instance, in the case of *Photuris* fireflies hunting *Photinus* males looking to mate, it was surely an adaptation of *Photuris* senders to produce the same signal as *Photinus* females in response to wanting a meal. Thus, on causal-informational teleosemantics the signal should mean *Photinus female looking to mate* OR *Photuris wanting a meal*. Infotel-semantics gets around this problem in the very same way biosemantics does, by appealing to the evolutionary success condition of the behaviour elicited in the receiver by the signal. In both the non-cooperative case of *Photuris* senders, and in the cooperative (or more cooperative) case of *Photinus* senders, the evolutionary success condition of the behaviour elicited in male *Photinus* receivers is that the sender is a *Photinus* female looking to mate. Thus, Infotel-semantics possesses the best of both worlds: the explanatory prowess of receiver-independent, correlational accounts of biological information, but without some of the problems that come along with them. Before moving out, however, it needs to be shown how infotel-semantics avoids the problem of non-cooperative communication facing biosemantics.

---

[25] See Carazo & Font (2010) for a strikingly similar account of the information content of signals, without the cooperation requirement.

Infotel-semantics, as articled by Shea (2007), essentially attaches to traditional, sender-receiver, biosemantics the idea that a signal must correlate with its content. That infotel-semantics builds on theoretical machinery of biosemantics is evident in the beginning of the above quote from Shea. Signals are intermediates in a system consisting of a sender and a receiver cooperating by means of a range of mediating signals. But, as I demonstrated in Section 2.3 (problem 2), this creates an issue when the interests of senders and receivers conflict. When interests conflict, senders and receivers do *not* cooperate by means of a range of mediating signals. What the sender wants is opposed to what the receiver wants. In response to this, my suggestion is that infotel-semantics be modified in the same way Stegmann modified traditional biosemantics. This would constitute only a minor change to infotel-semantics. None of the conditions (a) to (c) presented above need to be modified. Only the initial assumption that senders and receivers are cooperating needs to be jettisoned. But this does not change the fundamental recipe for assigning content: S must still correlate with its content; where this correlation must make sense of why receivers respond; and the content of the signal must be the evolutionary success condition of the receiver's response to S.

Take our *Photuris* imposter hunting *Photinus* males. S is the deceptive flash produced by *Photuris* hunters[26]. It carries the content *Photinus female looking to mate* because, (a), the flash raises the probability that it is produced by a Photinus female looking to mate. Of course, it raises the probability that it is produced by an imposter looking for meal too. However, (b) only the former correlation explains why receivers respond in species-typical fashion. Moreover, (c) only the former world-state (that the sender is a *Photinus* female looking to mate) constitutes the evolutionary success condition of male *Photinus* responses.

A couple of additional points about infotel-semantics are needed. As seen in condition (a), infotel-semantics takes it as constitutive of content that a signal S carries information about its content C. If there is no correlation between S and C in the present, then S doesn't represent C. One might worry about this. Does it entail that infotel-semantics has unfortunate verificationist consequences? It does not, because the correlation between S and C is a correlation between *types* not tokens. Tokens of S can occur in the absence of C. A sender can sometimes produce a leopard alarm call when in fact there is no leopard present. It just has to be that the leopard call *as a type* raises the probability that a leopard is present.

Nor does the content of the signal have to be the world-state the signal correlates *most reliably* with. A signal can correlate more reliable with a world-state(s)

[26] As mentioned previously, it might be individuated as the same signal as the honest one in virtue of having the same selected function: to attract *Photinus* males; and because it carries out this same function in virtue of the same physical properties *P1,…, Pn*.

that is not its content than with its content. This is because of clause (b) in infotel-semantics: namely, that an evolutionary explanation of the current existence of the signalling system must appeal to Ss having carried information about C for C to be the content. For a signal to have content C, it is not enough that the signal carried information about C. Over and above this, the fact that the signal carried information about C must explain the current existence of the signalling system: it must evolutionarily rationalise the response of the receiver. A leopard alarm will raise the probability that there is leopard present OR that the sender thinks there is a leopard present, more than it will raise the probability of the first disjunct alone. However, an explanation of the current existence of the vervet alarm communication system will not appeal to Ss having carried information about the sender being aroused. It will appeal only to Ss having carried information about the presence of a leopard.

This last point effectively circumvents one of Millikan's problems with correlational information; that it masquerades as a proto-semantic relation that exists independently of signal-*use*. Infotel-semantics, however, doesn't frame correlational information as determining content independently of use. It is open to the fact that a signal-type will correlate many world-states. When it comes to determining content, infotel-semantics picks out that world-state explaining the historical success of the behaviour elicited in receivers. Thus, while biosemantics maintains that content determination requires focusing on how signs are *used*, it goes too far, according to infotel-semantics, in holding that input conditions play *no* substantive content-determining role. Infotel-semantics gives a role to both input and output-based determinants of content. On the input side: those world-states that a signal carries correlational information about. On the output side: out of all those world-states, only those relevant to an explanation of the historical success of the receiver's response to the signal.

So far, I have outlined infotel-semantics as a plausible solution to the problem of circularity and explained how it can be made to handle non-cooperative communication. Now I offer some novel considerations regarding how infotel-semantics solves the two indeterminacy problems facing biosemantics remaining open from Section 2.3. First, the problem of omnipresent evolutionary success conditions. Recall the problem. There must have been oxygen in the atmosphere whenever the typical response to leopard alarms earned its keep. Gravity must also have been operating as normal. According to biosemantics, then, the content of the leopard alarm must be *leopard and oxygen in the atmosphere at sufficient levels and gravity operating normally*. However, infotel-semantics, because of its requirement that signals raise the probability of their contents, rules out omnipresent conditions. Nothing can raise the probability of conditions that are always present. Because the presence of oxygen in the atmosphere is omnipresent for all practical purposes, there is no greater chance of there being oxygen in the atmosphere given some other event (such as signal production). Thus, oxygen being in the atmosphere (as well any other omnipresent world-state) is eliminated as a candidate content.

Second, the problem of overly-specific evolutionary success conditions. Recall that the issue relates to the myriad ways in which the evolved response of the receiver could have failed. The receiver could have been unable to climb a tree successfully due to injury, or the receiver may have fallen out of the tree while climbing to safety due to grabbing a weak branch. There are numerous other world-states one can think up such that each must have been absent whenever the response to the signal earned its keep. Thus, there are numerous states that must be included in the content of the signal, according to biosemantics. In response, correlation can once again come to the rescue. A leopard alarm call doesn't raise the probability that receivers are well enough to climb a tree. The frequency of vervets well enough to climb a tree after a leopard alarm is tokened won't be greater than before a leopard alarm is tokened. Thus, such an overly-specific candidate content is ruled out. Likewise when it comes to grabbing weak branches. The frequency of vervets falling out of trees due to this problem after a leopard alarm is tokened won't be smaller than before the signal.

4. Conclusion

This chapter has examined different theories of content and assessed their suitability for animal signals. Drawing on the explanatory concerns of biologists, I claimed that the suitability of a theory of content for animal signals comes down to whether this content makes sense of why receivers respond to signals. I then examined: (1) causal-informational semantics, (2) input-orientated teleosemantics, (3) output-orientated teleosemantics ('biosemantics'), and finally (4) infotel-semantics, arguing that infotel-semantic content goes the farthest in meeting the desideratum. Moving forward, an infotel-semantic account of signal content will be in the background in Chapters 2 and 3, where I weigh into some conceptual debates over animal communication conceived informationally. As some in this literature have noted (Owren, Rendall & Ryan 2010), information proponents need a clearly-defined notion of signal content before engaging in some of these debates. I hope this chapter has contributed to laying that groundwork.

With a clearly-defined notion of "natural meaning" (Grice 1957) now established, the next chapter argues that signals cannot be accurately defined as traits that, uniquely, carry natural meaning. The resulting definition is too liberal, capturing most (if not all) co-adapted interactions between two organisms, as opposed to those interactions that are conceived as communicative. I argue that this is problematic for all currently-available definitions, as an informational definition was intended as an improvement over definitions eschewing information, which are themselves too liberal. Chapter 3 then weighs more heavily into the 'information vs influence' debate. There I argue that, while it makes sense to attribute natural meaning to signals, we must make two concomitant distinctions: between 'proximate' vs 'ultimate' explanatory roles of this meaning; and concomitantly, between two different ways this meaning is 'derived' by receivers from signals. The first way is cognitively modest, and is probably the most widespread way receivers derive information from signals. Here, the natural meaning of signals plays merely an ultimately explanatory role. The second way receivers derive natural meaning from signals is cognitively richer, and plausibly less widespread. Here, the information contained in a signal plays a more-or-less proximate explanatory role. These distinctions are advanced in the hope of alleviating the mechanistic concern of information sceptics, discussed in the preface. The concern, if you recall, is that explaining receiver responses in terms of derived information often inflates the cognitive sophistication of receivers.

*   *   *

Chapter 2. *Can we uniquely define animal communication in terms of the exchange of natural meaning, or does such an exchange occur between all co-adapted interactions? Perhaps signals, uniquely, are arbitrary? What is arbitrariness and what is the relationship between arbitrariness and signalhood? Do any criteria cleanly separate communicative co-adapted interactions from non-communicative ones?*

Chapter 2

Problems for Defining Signalhood in Terms of Natural Meaning

Biologists classify some, but by no means all, between-organism interactions as 'communication' (or 'signalling'). Yet, it is contentious exactly what features an interaction must possess to count as communication, as opposed to, say, a case of one organism *coercing* another, or alternatively one organism inadvertently producing a *cue* that can be utilised by another (Maynard-Smith & Harper 2003; Scott-Phillips 2008; Carazo & Font 2010; Scarantino 2013). In short, what makes only some behaviours or morphological traits 'signals'? In this Chapter I argue that no currently-available definition, particularly a recently-proposed informational definition, is extensionally adequate. This chapter is an exercise in the philosophy of nature just as much as the philosophy of science. Its focus is on nature itself, and how nature is to be carved at its 'joints', just as much as it is about scientific explanation.

1. Communication as co-adapted influence?

Before the rise of individualist and gene-centric views of evolution in the 1960s and 70s, the idea that natural selection adapts senders to faithfully 'inform'[27] receivers was commonplace. However, after the demise of 'naïve' group selection, this idea became somewhat controversial (Dawkins & Krebs 1978). The worry was that natural selection, operating at the level of the individual organism, will very often adapt senders to manipulate or influence receivers, not 'inform' them. This is because the fitness interests of senders and receivers commonly diverge to varying degrees. For example, a relatively unfit male frog will not be selected for indicating his sub-par genetic quality to a female looking to mate. Rather, he will be selected for doing whatever he can, causally speaking, to influence the female to mate with him, regardless of the interests of the female.

[27] 'Informing' is meant in a loose sense here: as causing a response in a receiver that is adaptive for the receiver, given the state of the world.

Moreover, taken as a defining feature of signalhood, 'information transmission' is too liberal. This is because various other behaviours and morphological features of organisms, aside from signals proper, transmit information. Consider a stag (I1) coming up against another stag (I2) in a contest over a contested resource. After assessing the size of I2, I1 decides not to engage with I2. Here I2's size is a source of information that I1 utilises in deciding how to act (or, in this case, how not to). But it isn't a signal. Rather, it is a cue (Maynard-Smith & Harper 2003). Thus, a definition of a signal that defines a signal as any trait that transmits information to another organism fails. It is too liberal: it can't exclude cues.

In response to these problems with a definition of communication couched in terms of 'information transmission', some took a different approach entirely. They avoided reference to information transmission and instead formulated a definition based solely on sender *influence*. On this definition, signals are traits adapted for causally influencing a receiver (Maynard Smith & Harper 1995). This influence-based definition was able to differentiate signals from cues. The larger size of stag I2 has a causal effect on the smaller stag I1, but it is not a trait that was selected for influencing I1. I2's size has a causal effect on I1, but from the point of view of I2, this effect is inadvertent. Size (in itself) is not an adaptation for intimidating rivals[28].

Compare this 'inadvertent' and thus non-communicative sort of behavioural influence with a genuine signal. The elongated tail of the male barn swallow is a signal of high genetic quality (Moller & de Lope 1994). Like in the case of stag size, an elongated barn swallow tail elicits a behavioural effect in a conspecific (namely, sexual receptivity in a female). But unlike in the case of stag size, an elongated tail has the selected function of eliciting a behavioural response in a conspecific. An elongated tail is an adaptation for eliciting sexual receptivity in female barn swallows. In contrast, the size of the larger stag is not an adaptation for eliciting a disengagement response in smaller stags. Accordingly, Maynard-Smith & Harper (1995, p. 306) defined, in non-informational terms, a signal as any "action or structure that increases the fitness of an individual by altering the behaviour of other organisms detecting it, and that has characteristics that have evolved because they have that effect".

However, this solely influence-based definition was itself extensionally inadequate. It, too, was overly liberal. It could not differentiate signals from acts of coercion. The basic problem is that many kinds of traits are adaptations for influencing other organisms. A stronger stag pushing a weaker stag in a confrontation over a contested resource is a behavioural adaptation of the stronger stag that is selected for influencing the weaker stag. But it is not a signal. As a result, Maynard-Smith & Harper (2003) strengthened their earlier definition of a signal. As before, a

[28] Of course, behaviours that are adapted for advertising or accentuating one's size might be signals. But here we are not dealing with such adaptations: only size *as such*.

signal must be an adaptation (of the sender) designed to influence a receiver. But they added the requirement that the elicited response in the receiver must also be an adaptation (of the receiver): "We define a 'signal' as any act or structure which alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved" (Maynard-Smith & Harper 2003, p. 3).

In subsequent analyses, Scott-Phillips (2008) and Scarantino (2013) noted that there is some ambiguity in how we interpret Maynard-Smith & Harper's (MSH's) more restrictive, influence-based, definition. The ambiguity relates to how we interpret the new requirement, related to the adapted nature of the receiver's response. We can interpret it in one of two ways: (a) the receiver's response to a signal must be an adaptation, but not necessarily an adaptation to the signal; (b) the receiver's response must be an adaptation, specifically to the signal. Following Scarantino (2013), we can call the first way of interpreting the requirement the *generalised response interpretation* and the second way the *specialised response interpretation*.

Scott-Phillips (2008) argued that an adequate influence-based definition of communication must adopt the specialised response interpretation. That is, the sender's act or structure must be "effective because the effect (the response) has evolved *to be affected by the act or structure*" (Scott-Phillips 2008, p. 388, my italics). The response should not merely be adapted, it must also "be adapted to fulfil its half of the communicative dynamic" (Scott-Phillips 2008, p. 388).

According to Scarantino (2013), MSH (2003) had in mind the specialised, as opposed to general, response interpretation. This is shown by the fact that MSH exclude camouflage from being a signal. Camouflage can't be excluded as a signal based on the generalised response interpretation. It can only be excluded based on the specialised response interpretation. A predator failing to perceive a camouflaged prey *as prey* is due to evolved properties of the predator's sensory organs. Thus, the prey (the would-be 'sender') elicits an effect in a would-be 'receiver' (the predator passing over the prey), where this effect relies on the adapted sensory organs of the 'receiver'. Moreover, eliciting this effect in the 'receiver' is an adaptation of the 'sender'. So, camouflage is a signal under the generalised response interpretation of MSH's (2003) definition. But if we operate with the specialised response interpretation, camouflage can be excluded as a signal. This is because the predator's response to the camouflage is not an adaptation to the signal specifically. Rather, the response is an adaptation to all kinds of environmental stimuli[29].

In sum, at this point in the historical dialectic it appeared a definition of signalhood was prevailing that: (1) eschewed information (to differentiate signals and

[29] This assumes the predator has not evolved specific perceptual mechanisms in response to the camouflage which are fooled, in an evolutionary arms race.

cues), and (2) stressed the co-adapted nature of communication (to differentiate signals and coercive acts). Scott-Phillips (2008, p. 388), making explicit the specialised response interpretation of MSH's (2003) definition, defined a signal in the following terms:

> A signal is any act or structure that (i) affects the behaviour of other organisms; (ii) evolved because of those effects; and (iii) which is effective because the effect (the response) has evolved to be affected by the act or structure.

For present purposes, let's call this definition of a signal a *co-adapted influence* definition. Components (i) and (ii) demand that communication is a matter of one organism (the sender) being adapted to causally influence another organism using a morphological structure or a behaviour. Component (iii) demands that this adapted influence be a two-way street: receivers must have evolved to be affected by the relevant behaviour or morphological structure.

1.1 Recent criticism of communication as co-adapted influence

Recently, a co-adapted influence definition of communication (i.e. MSH (2003), Scott-Phillips 2008)) has come under fire. According to Scarantino (2013), whose argument will be the focus in what follows, a co-adapted influence definition is itself too liberal (even on the specialised response interpretation). It includes within its extension various co-adapted interactions between two organisms that are not genuine cases of communication: namely certain kinds of coercive and reciprocal interactions. Scarantino (2010; 2013)[30] argues that only by adding an informational component to a co-adapted influence definition can the definition be salvaged. First, consider the following kind of coercive interaction:

> A weaker stag keeps being pushed backwards by a stronger stag. This goes on for a while, until the weaker stag adopts behavioural responses to the negative happening to which he is being subjected, namely being pushed. In the short term, the response may be to give up on the contested resource. In the medium and long term, the response may be to accept a more submissive position in the social hierarchy, and avoid further confrontations with the stronger stag. (Scarantino 2013, p. 75).

We can safely assume that the stronger stag's pushing behaviour is an adaptation: it was selected for influencing the weaker stag because this was a way for the stronger stag to attain the contested resource, to maintain its place in the social hierarchy, and

[30] Carazo & Font (2010) also argue for a hybrid definition of communication invoking both (co-adapted) influence and information transmission. Their reasons are somewhat different to those of Scarantino, but the same upshot of my discussion of Scarantino's definition applies equally to Carazo & Font's. The latter (2010, p. 663) propose that a signal is "any act or structure that (i) affects the behaviour of other organisms; (ii) evolved (or is maintained) because of those effects; (iii) is effective because it transfers… information to receivers. Their account of information is very similar to the one I advanced in Chapter 1, requiring correlation and teleology on the side of the receiver without cooperation.

to avoid future conflicts. Equally, we can safely assume that the response of the weaker stag is an adaptation (for the weaker stag): giving up on the contested resource, when faced with a stronger competitor, was selected for in the face of potential injury from a prolonged contest. Also, accepting a more submissive position in the social hierarchy was selected in the face of costly future conflicts likely to be lost.

Furthermore, Scarantino argues that the weaker stag's responses evolved specifically to the act of being coerced by the stronger stag: "It seems reasonable to posit evolutionary pressures selecting in favour of organisms that take being coerced as a sign of a competitor's superior strength, and change their future behaviour towards that stronger competitor accordingly" (Scarantino 2013, p. 76).

Thus, coercive behaviours with this kind of structure count as signals on Scott-Phillips (2008) co-adapted influence definition of communication: "They alter what organisms do as a result of being coerced, they evolved at least in part because of that effect, and they are effective because the receiver's response has also evolved" (Scarantino 2013, p. 76). However, according to Scarantino (2013, p. 76), "any definition of animal signalling that counts coercive behaviours as signals is too broad". Second, consider the following kind of reciprocal interaction:

> In cases of reciprocity, actors take turns in benefitting each other. In each interaction, the giver produces a positive happening for the receiver (e.g. being groomed)… grooming behaviours alter the behaviour of another individual, and presumably evolved to influence the recipient's behaviour towards reciprocation, because this is the main expected payoff of an otherwise costly activity. Finally, grooming behaviours appear to be effective because the receiver's response to the behaviour has also evolved. This is true both in the sense that the response depends on evolved properties of the brain and sensory system of the groomed, and in the sense that the response – reciprocating a positive happening – evolved specifically as a response to earlier grooming behaviours[31]. (Scarantino 2013, p. 76).

Thus, certain kinds of both coercive interactions and reciprocal interactions are categorised as cases of communication by a co-adapted influence definition. The point can be expanded to a more general one. Any time we have co-adapted influence, we will have communication, on the definition currently on trial. As a result, Scarantino presents a new definition of communication that sets out to be more restrictive.

2. Communication as 'Information-Mediated Influence'?

Scarantino's proposed definition is hybrid in nature, incorporating both co-adapted influence and information transmission. The co-adapted nature of communication remains definitive: the sender's act of influence must be an adaptation of the sender.

[31] When it comes to great apes the response to an initial act of grooming might be produced by general-learning mechanisms, but it can still be said that the response itself arose, over ontogenetic time, in response to the initial act as opposed to any kind of stimuli.

Likewise, the receiver's response to the act of influence must be an adaption of the receiver (to the signal). But co-adapted influence is not enough.

Despite being instances of co-adapted influence, Scarantino argues that pushing and grooming are not signals. This is because "a signal must be specialised for *influencing recipients by carrying information to them*" (Scarantino 2013, p. 77, author's italics). Pushing and grooming are "instead specialised for influencing recipients by other means, namely the production of, respectively, negative and positive happenings through *mechanical interactions* (Scarantino 2013, p. 78, my italics). But what, exactly, is the nature of an act of influence specialised for causally influencing a recipient by carrying information to them? Answering this question involves saying briefly what is meant by carrying 'information'; and then making sense of how information, thus conceived, plays a causal-explanatory role in behaviour.

First, Scarantino (2013, pp. 64-66) characterises information as "predictive information"[32]. A signal (S) can be understood as carrying predictive information in virtue of changing the probability (P) of some state(s) of the world (W). S carries information about W if and only if P(W/S) does not equal P(W). If P(W/S) is greater than P(W), then S has raised the probability that W obtains. When S raises the probability of W, it carries information about W obtaining. Conversely, if P(W/S) is lower than P(W), then S has lowered the probability that W obtains. When S lowers to probability of W, it carries information about W not obtaining.

To narrow down the actual content of the signal we can simply add teleology on the side of the receiver, in line with Chapter 1. This will not change the predictive information carried by a signal, nor the conclusion of this chapter. Thus, in what follows I will speak of the natural meaning carried by a signal, where this can be understood as entirely compatible with predictive information.

Now we come to the question of the causal-explanatory role of the natural meaning carried by a signal. More specifically, what does it mean for a receiver to be causally influenced by a signal in virtue of the latter's natural meaning, as opposed to the signal's mechanical properties? Answering this question is not obvious. Indeed, the information vs influence debate has recently arisen surrounding whether the idea can be made coherent. Information-sceptics lament what they see as a misguided turning away from the intrinsic physical properties of signals responsible for triggering responses in receivers, towards the idea that information is a separate entity 'carried' by or 'contained' in signals which does the response triggering (Rendall et al. 2009; Owren et al. 2010). Sceptics argue that this buys into the 'conduit metaphor' of communication (Reddy 1979; Blackburn 2007).

---

[32] It is more or less the same as the notion of correlational information advanced by Shea (2007) and Skyrms (2010). Scarantino gives a richer and more detailed analysis of this kind of information in his (2015).

The conduit metaphor of communication is usually discussed in the context of linguistic communication: specifically, as being a misguided way of describing the actual process of successful communication. We commonly speak of getting one's ideas *across* in a conversation, or of an utterance *containing* many good ideas (etc.). But in reality, linguistic communication does not actually involve 'meaning', 'content' or 'information' being packaged into an utterance by a sender, where this content is subsequently unpacked from the utterance by the receiver after transmission. Although there is some controversy over what, exactly, content is, one thing it definitely isn't is a substance that placed into an utterance and then transmitted from point A to point B. When it comes to the content of linguistic utterances, one influential view is that content is a function or rule which tells you what a word or sentence would refer to, if the world was in a certain way. Another way of putting this is that contents are functions from possible worlds to truth-values (Nolan 2005)[33].

Now, it is vital to note that both Scarantino's (2013; 2015) account of predictive information, and the account of natural meaning defended in Chapter 1, are both non-metaphorical and non-conduit based. They both denotes a well-defined relation between signal and world state, as opposed to a reified substance that is first 'packaged' into a signal by a sender and subsequently 'unpacked' by a receiver in the process of communication. Still, there remains the issue of making sense of how natural meaning, as distinct from a signal's 'mechanical' or intrinsic properties, explains receiver behaviour.

In Section 2.1 I suggest how the natural meaning 'carried' by a signal might explain the behaviour of a receiver influenced by the signal. However, in Section 2.2 I argue that, understood in this way, natural meaning is present in most, if not all, co-adapted interactions. If this is correct, then defining communication informational terms will be too liberal.

2.1 THE CAUSAL-EXPLANATORY ROLE OF NATURAL INFORMATION

What does it mean to say that a receiver is causally influenced by the natural meaning of a signal? In other words, what causal-explanatory role does this meaning play in understanding a receiver's respond to such a signal? In proposing an answer to this question, I will draw on some philosophical machinery belonging to Fred Dretske (1994). More specifically, his distinction between a *triggering* cause of behaviour, on the one hand, and a *structuring* cause of behaviour, on the other.

---

[33] There is then the further foundational question of what facts make it the case that an utterance has this or that truth-condition. One influential view comes from Grice (1957), and takes the content of utterances to be grounded by the content of mental states. Roughly speaking, my utterance means what it does because of my goal in producing it; and you understand my utterance by (accurately) inferring this goal of mine.

Dretske (1994) himself presented the distinction in the context of someone moving a cursor on a computer screen by pressing the 'backspace' key on the keyboard. Pressure on the key is a triggering cause of the cursor's movement. But there is also a different sense in which the particular wiring of the keyboard, for example, and the program of the word-processor, are causally responsible for the cursor's movement: they create the "standing conditions" within which backspace presses trigger cursor movements. It is in this sense that the particular wiring of the keyboard and the program of the word-processor are each structuring causes of the cursor's movement. They both cause (2nd order) the conditions under which, in turn, backspace presses cause (1st order) cursor movements.

Turning back to communication, we can understand the natural meaning carried by a signal to be a structuring cause of receiver behaviour. In contrast, the mechanical features of a signal can be understood as triggering causes of receiver behaviour. For example, vervet monkeys run up trees when they sense a kind of acoustic pattern emitted by a conspecific: namely a leopard alarm call (Seyfarth et al. 1980). The token acoustic pattern can be regarded as the triggering cause of the vervet's tree-climbing behaviour. However, a learning history in which vervet alarm calls raised the probability of the presence of a leopard also explains the tree-climbing response of the receiver: but as structuring cause as opposed to triggering cause.

To adequately explain why receiver vervets now run up trees in response to leopard alarms, we must note the historical correlation between leopard alarms, on the one hand, and certain facts about the world that 'rationalise' the response elicited by such calls in the receiver, on the other: namely that a leopard was present. This historical correlation between leopard alarms and actual leopards is causally responsible (2nd order) for the conditions within which leopard alarms now cause (1st order) tree-climbing behaviour[34]. This is like the way in which the particular wiring of the computer's keyboard and the word-processor's program are causally responsible (2nd order) for the conditions within which backspace presses now cause (1st order) cursor movements.

The proximate/ultimate distinction in biology is also useful for illuminating the causal-explanatory role of natural meaning[35]. Following Mayr (1961, p. 1503),

[34] If leopard alarms did not correlate with the presence of a leopard over the learning period, receivers wouldn't respond to them by climbing trees.

[35] There has been significant controversy about this distinction in recent years (e.g. Ariew 2003, Laland et al. 2011, Dickins & Barton 2013, Scholl & Pigliucci 2015). Despite this, even critics of the distinction such as Laland et al. (2011) are not entirely dismissive, acknowledging that it is useful in cases where reciprocal causation between development and natural selection are not paramount: where the organism has been shaped unidirectionally by selection to match the external environment. However, I side with Dickins & Barton (2013) who argue that the proximate-ultimate distinction in no way blinds us to developmental influences on evolution, when these are in fact paramount, and that the proximate-ultimate distinction is in fact needed to highlight such influence in the first place. Development, understood (as with Mayr) as a proximate process, is of course influenced by evolution but can also influence evolution.

proximate causes "govern the responses of the individual (and his organs) to immediate factors of the environment" while ultimate causes "are responsible for the evolution of the particular DNA code of information with which every individual of every species is endowed". Proximate causes are immediate and mechanical, while ultimate causes are historical and made sense of in terms of success (either over phylogeny or over ontogeny[36]). Mayr's famous example is of bird migration. There are two kinds of causal explanation for why a bird migrates for winter. The first, proximate explanation, relates to the bird's present physiology and the influence of day length on the concentration of hormones in the bird's brain. The second, ultimate explanation, relates to the historical fact that migrating for the winter allowed the bird's ancestors to avoid starvation. Similarly, there are two kinds of causal explanation for why a vervet receiver flees up a tree in response to a leopard alarm. The first, proximate explanation, refers to mechanistic factors such as the acoustic form of the signal and its interaction with the nervous system of the receiver. The second, ultimate explanation, refers to the historical fact that the leopard alarm correlated with the presence of a leopard over a learning period. These are two different, but of course complementary, reasons for why vervets climb trees in response to leopard alarms.

Both the structuring cause vs triggering cause distinction, and the proximate vs ultimate distinction, allow us to get clear on the causal-explanatory role of natural meaning in communication. I hope that elucidating the causal-explanatory role of predictive information will also render the notion of *information* more palatable for those sceptical of the utility of informational notions in explaining animal communication (e.g. Owren et al. 2010; Rendall et al. 2009)[37]. Often, natural meaning is a 2nd order cause of receiver behaviour while 'mechanical' factors are 1st order causes. Alternatively, natural meaning is often an ultimate cause of receiver behaviour while 'mechanical' factors are proximate causes. However, I argue in Section 2.2 that, understood in this way, natural meaning plays a causal-explanatory role in many, if not most, co-adapted interactions. If this is true, then an information-mediated influence definition adds very little to a definition of communication stressing merely co-adapted influence (e.g. Scott-Phillips 2008).

2.2 A PROBLEM FOR AN INFORMATION-MEDIATED INFLUENCE DEFINITION: INFORMATION ISN'T SPECIAL

[36] Mayr had phylogeny firmly in mind but there is no reason why an ultimate explanation for some behaviour cannot appeal to success over ontogenetic history.
[37] I have much more to say on this issue in the next chapter.

Consider again the case of the two stags pushing each other in a confrontation over a contested resource. Scarantino argues, plausibly, that this is a co-adapted interaction: I1's act of pushing is an adaptation for I1, and I2's responses (i.e. giving up the immediate contest, and assuming a lower place in the social hierarchy) are adaptations (for I2) specifically to the act of repeatedly being pushed backwards. Granting the co-adapted nature of the interaction, let us briefly investigate the details surrounding the evolution of the I2's responses. We will see that natural meaning has just as much of a (structuring) causal role in explaining I2's behaviour as it does in genuine cases of communication.

Remember that, in the stag contest, I2's response to being pushed backwards is two-folded. In the short term I2's response is to give up on the contested resource. In the long run I2's response is to assume a lower place in the social hierarchy. Crucially, making sense of why these two responses evolved in the face of losing a pushing contest (i.e. being repeatedly pushed backwards by a conspecific stag) involves noting a historical correlation between the stimuli eliciting the two responses in I2 (i.e. the pushing), on the one hand, and certain facts about the world that 'rationalise' these responses in I2, on the other: in this case the fact *that the stimuli were produced by a stronger stag*. Only when I2 being pushed by I1 correlated with I1 (the pusher) being stronger than I2 did it make adaptive sense for I2 to choose to give up on the contest, in the short term, and to assume a lower place in the social hierarchy, in the long term. Thus, I1 getting I2 to submit (in both the short and long term) literally carries natural meaning about I1 being stronger and thus likely to inflict potential further damage on I2, win future conflicts against I2, and so on. And because, in addition to carrying this meaning the interaction fits all the relevant criteria for being a co-adapted interaction, I1 pushing I2 backwards turns out to be a signal on an information-mediated influence definition of communication.

Consider grooming, as well. Here, individual 1 (I1) grooms individual 2 (I2), thereby eliciting a response from I2 (namely, I2 becoming more disposed to engage in future cooperation with I1). Both I1's act and I2's response are adaptations, respectively, for I1 and I2. As before, when we inquire into the details surrounding the evolution or development of I2's response in this co-adapted interaction, we will be forced to note a historical correlation between the stimuli that causes I2 to become more benevolently inclined towards I1 (i.e. the initial act of grooming), on the one hand, and certain facts about the world that 'rationalise' this response in I2, on the other: in this case *that I1 is less likely to attack I2 upon I2 reciprocating* (among other facts). Thus, I1's act of grooming literally carries natural meaning about I1 being more likely to accept future personal contact from I2. And, again, because in addition to carrying this information, the interaction fits all the relevant criteria for being a co-adapted interaction, I1 grooming I2 is a signal, according to an information-mediated influence definition of communication.

The problem can be stated in a general way. Firstly, we have two sets of cases, C1 and C2; where both sets of cases are comprised of co-adapted, between-organism interactions with the following structure: An individual (I1) influences another individual (I2), where the selection/learning history of I2's response to I1's initial act of influence crucially involves a probability-raising correlation between this act of influence and some world-state that 'rationalises' the response elicited in I2. Secondly, however, we require that only in C1 cases was this natural meaning 'causally relevant' or 'exploited' by receivers, while in C2 cases it was not: allowing us to classify C1 cases, but not C2 cases, as 'communicative'. As things stand, I do not see a way in which to do this. In both sets of cases, I1's act of influence carries natural meaning, and this information is a structuring cause of I2's evolved response to I1's act of influence. Vervet alarm calls raise the probability that a leopard is present. Similarly, a monkey grooming a conspecific raises the probability that it is not going to attack the conspecific when the latter reciprocates.

A possible reply at this point in the argument might be this: 'OK, even stag pushing contests and acts of grooming 'transmit information', but not about external states of affairs.' The idea behind the reply is that the information transmitted by a stronger stag pushing a weaker stag backwards and winning a contest is about the stronger stag (i.e. literally that he is the stronger competitor). But the information transmitted by, say, a vervet leopard alarm call, is about the presence of a leopard, which is a world-state external to the sender. However, it is obvious why this reply fails. Most cases of signalling involve information transmitted about the sender, as opposed to the external environment. Despite their paradigmatic status in much of the communication literature, alarm calls are the exception[38]. The peacock's famous tale, indicating high genetic quality, is but one example of a signal transmitting information about the sender. A large and impressive tail raises the probability that the male sender is of good stock, and this is the state that 'rationalises' the adapted response of the female receiver.

The moral is this: conceived of appropriately, i.e. as a structuring cause of receiver behaviour, natural meaning is 'transmitted' in all sorts of co-adapted interactions. This should be no surprise, really. For presumably most, if not all, co-adapted interactions must occur in the right circumstances. There will always be a world-state that 'rationalises' the recipient's response such that the response needs to occur in that state of the world (even if this be internal to the sender), which will also mean the 'sender's' eliciting act needs to correlate with the rationalising world-state.

[38] Why this obtains, when models (Skyrms 2010) show us that signalling about the external environment should readily emerge, is raised by Sterelny (2017).

This largely completes my analysis of an information-mediated influence definition of communication. The moral I would like to draw is the following. If an information-mediated influenced definition of communication is to be salvaged, it would have to be shown that the natural meaning carried by signals can be characterised in a way that is different from the natural meaning pervasive among co-adapted interactions more generally. In the absence of such an account, an information-mediated influence definition seems to suffer from the same problem as a co-adapted influence definition: both are too liberal.

However, before ending I would like to propose one way of modifying an information-mediated definition to make it less liberal. The modification involves augmenting an information-mediated influence definition with the notion of 'arbitrariness' to rule out co-adapted interactions like stag pushing and grooming behaviours. However, despite being able to rule out the latter, I will argue that this proposal is unwise. It swings things too far in the other direction, and makes a definition of communication too restrictive. This is because many kinds of signals are not arbitrary.

3. Adding an Arbitrariness Feature?

Take the famous alarm calls of the vervet monkey. The 'arbitrary' nature of these calls was originally seen as one of their most exciting features: one suggesting possible continuity with human language (Seyfarth et al. 1980). More specifically, the link between (a) the physical structure of each alarm call, on the one hand, and (b) the function of each alarm call, on the other, was found to be arbitrary. Put colloquially, vervet receivers wouldn't have cared if the physical structure of the actual leopard alarm took the structure of what is now the eagle alarm (and vice versa). What is important is simply that the leopard alarm, whatever physical structure it takes, be distinguishable from an eagle alarm and a snake alarm, and that it correlate more strongly with the presence of leopards as opposed to eagles or snakes. But this can occur across multiple variations in physical signal structure: so long as each call can be perceptually distinguished from each other and from any other signals vervets have that are unrelated to predator avoidance[39].

Understood this way, a signal can be arbitrary in two senses: the first 'all-or-nothing'; the second contrastive. In the first sense, a signal is arbitrary if at least one

[39] This is not to say that vervet alarm calls are completely arbitrary, in the sense I am beginning to outline. As Owren et al. (2010) rightly note, the physical form of alarm calls is constrained to some degree. They must allow for detection and localization. Also, they must prime listeners by engaging low-level attention and arousal mechanisms. However, there seems to be no necessary connection between the physical form of a leopard call, on the one hand, and behavioural responses that are adaptive in the presence of leopards as opposed to eagles or snakes, on the other. In other words, there is nothing special about the physical form of leopard calls, qua signals designating leopards, as opposed to that of eagle or snake calls.

alternative physical structure could have been used in place of the actual one. In the contrastive sense, a signal is *more* arbitrary the greater the range of alternative physical structures that could have been used in place of the actual one. But how, exactly, do we determine whether such-and-such alternative structure *could have been used* in place of the actual one? This is a complex issue. However, there are plausible constraints. An obvious constraint is perceptual discriminability. Signals must be detectible by receivers. Second, producing the alternative physical structure must have been within the behavioural repertoire of senders at the time of signal creation. This rules out alternative structure that would have required senders to possess novel perceptual-motor or cognitive capacities, for instance. A third potential constraint I would suggest, albeit tentatively, is that some members of the population must have been disposed, at the time of signal creation, to produce the alternative structure in response to the same condition others were disposed to respond to with the actual structure.

Accordingly, vervet alarm calls are relatively arbitrary in the contrastive sense because each type of call was obviously within the behavioural repertoire of senders at the time of signal creation; and also because, plausibly, some members of the population were disposed at that time to produce each kind of acoustic structure for each of a leopard, eagle or snake. Further, it seems plausible to think that the range of alternative acoustic structures that could have been used in place of the actual one - for each kind of alarm call - extends beyond the three kinds of structures that are actually used for alarm calling. But now compare vervet alarm calls with linguistic signs. While there are alternative physical structures the leopard alarm might have taken, the word 'l-e-o-p-a-r-d' is probably even more arbitrary. A much greater range of physical structures could have been substituted for the word 'l-e-o-p-a-r-d' at the time of initial signal development.

Now, arbitrariness is a property of interactions held by some ethologists and philosophers to be prototypically 'communicative'[40]. According to one prima facie plausible way of thinking about signals as distinct from other kinds of behaviours or phenotypes, the former involve "a distinctive role for relations of involvement between [signals] and other things" (Godfrey-Smith 2014, p. 83). The idea is that a paradigm signal, as opposed to a non-signal, brings about its effects *conventionally* as opposed to via its intrinsic properties. Skyrms (2010, p. 7) characterises as arbitrary (he uses the label 'conventional') signals that "are not endowed with any intrinsic meaning". "If we start with a pair of sender and receiver strategies, and switch the messages around the same way in both, we get the same payoffs" (Skyrms 2010, p. 8).

[40] See Harms 2004, Skyrms 2010; see also Wheeler & Fischer 2012 for a discussion of 'functional reference' in the ethology of communication, which made much of the arbitrary nature of communication in species like vervet monkeys.

"This fundamental symmetry is what makes Lewis signalling games a model in which the meaning of signals is *purely conventional*" (Skyrms 2010, p. 8).

Consider the mating signals of a female firefly, given in response to the flash of a male, which, when sent by the female, functions to attract the male (Lewis & Cratsley 2008). The flash brings about its effect of attracting a male by 'standing in' for a female firefly being ready to mate. Another way a female could get a male to mate with her might be to approach a male and perform some firefly equivalent of an enticing dance in front of him. Comparatively speaking this would not be a behaviour that brings about its effects conventionally. Instead, this behaviour is efficacious in virtue of its intrinsic properties. Likewise, there are two ways in which someone could come to believe that I'm hungry. I could utter to them the statement 'I'm hungry'. Alternatively, I could let them hear my stomach rumbling. There are also two ways an infant could come to know that it's a good time to feed. Mother could give off a signal that stands for the fact that she is ready to feed. Alternatively, mother could simply initiate feeding behaviour by exposing her nipple in direct view of the infant. In each case there is a difference grounded in the arbitrariness of the 'communicative' act, compared with the non-arbitrariness of the 'non-communicative' act. On this view, signals are ultimately efficacious because of sender-receiver co-adaptation as opposed to the intrinsic features of the signal.

This is not to say that in real time signals don't bring about their effects due to their intrinsic properties. Signals don't operate by magic, mechanistically speaking. Instead, the point is related to notion of arbitrariness outlined above. When it comes to a signal as opposed to a non-signal, a greater range of alternative physical structures could have been used at the time of initial signal development. But now this point has a bit more context. It is because receivers 'choose' to respond to signals, unlike some other kinds of stimuli, that the physical form of signals is relatively arbitrary. Of course, this 'choice' is often made by natural selection as opposed to the receiving organism itself. When natural selection stabilises the communication system, receivers attend to signals because signal-effects are adaptively beneficial. While the means are different, the comparison with human language is clear. Counterfactually speaking, I attend to the utterance "the building is on fire" not because of its intrinsic properties but because of what it conventionally stands for, given my desire to stay alive unmolested by hot flame. This contrasts with firefly mating signals, where it is natural selection that 'makes the choice' to keep receivers attending to signals.

With is in mind, we can now see how the notion of arbitrariness might be used by an information-mediated influence definition to exclude co-adapted interactions like stag contests and grooming. Compare a paradigmatic signal like a leopard alarm call with a stag pushing another stag in resource contest. One might think that the relatively arbitrary nature of the leopard alarm, in contrast to stag pushes, renders the former but not the latter co-adapted interaction 'communicative'. Although difficult to quantify precisely, it seems plausible to think that more alternative physical structures could have been used as the leopard alarm than as a trait for getting competitors to yield a contested resource. Presumably, getting a conspecific to give up a resource in a fight can only be done in limited number of ways given the physical structure of the world. It involves somehow stopping an opponent from getting something you want, via physical coercion. By way of contrast, a comparatively large range of possible vocal calls (with the right physical features allowing for detection, localisation, and eliciting emotion arousal) could presumably be used by vervets to correlate with leopards.

A similar comparison might be made between the leopard alarm, on the one hand, and 'non-communicative' grooming interactions, on the other. Compared to the range of physical structures that could play the leopard alarm, those that cause conspecifics to be more inclined to reciprocate with you in the future are probably more restricted (especially if you are a chimpanzee). You must do something which confers a benefit on a conspecific, which is plausibly more restricted than producing an acoustic call of some kind that correlates with leopards.

An information-mediated influence definition of animal communication bolstered by the notion of arbitrariness would look something like the following. While quite a mouthful, it captures how arbitrariness could be combined with an information-mediated influence definition to rule out instances of co-adapted influence like pushing contests and grooming:

> Phenotype P of a sender S is a signal when P has the selected function to trigger a certain response in a receiver R, where R's response is co-adapted specifically to the receipt of P, where P carries natural meaning which is a structuring cause of R's response, and where, finally, the link between the physical form of P and its function is (relatively) arbitrary.

3.1 ARBITRARINESS IS TOO RESTRICTIVE

While adding an arbitrariness criterion to an information-mediated influence definition might seem like a promising way of cordoning off communication from other co-adapted interactions, there is a problem. The problem is that many cases of communication aren't all that arbitrary. Worse, certain paradigm communication systems, such as the famous waggle dance of the honeybee, don't seem to be all that arbitrary. Take the way in which an incoming bee signals the direction that nectar-seeking conspecifics should fly upon leaving the hive. It orientates its dance relative to the vertical of the hive. The angle of the dance off the vertical of the hive maps onto the angle relative to the sun's azimuth receiver bees need to take to find food. For instance, a dance at 90 degrees from the vertical of the hive communicates nectar at 90 degrees from the sun's azimuth (at the time of signalling).

It is probably not the case that the physical structure of the actual signal could have been different, while nevertheless causing the same response. For the fact that bees presently orientate their dance relative to the vertical of the hive looks to be a product of evolutionary ritualization. At some earlier stage of bee evolution, a worker would fly back to the hive to deliver nectar and then ready itself to fly back to the same location from which it got its latest batch of nectar. As it prepared to exit the hive, it orientated its body in the appropriate direction. Thus, the direction component of the dance was probably originally a *cue* which provided other bees with information about the location of nectar (von Frisch 1967).

Similarly, "indices of quality" (Maynard-Smith & Harper 2003) are a relatively non-arbitrary class of signal because of diverging interests between senders and receivers. For example, male toads produce aggressive vocal signals designed to deter other males in disputes over access to females (Davies & Halliday 1978). Such vocalisations are nomically correlated with body size, and thus fighting ability. Low-pitched vocalisations, by their very nature, tend to be produced by larger toads, and hence are more effective at settling disputes in the sender's favour. Selection acting on male responses has effectively imposed an 'honesty' demand upon the signal type. This honesty demand drastically cuts down on the space of possible alternative physical structures that could function to intimidate rivals.

Yet more signals that an information-mediated influence definition (augmented with an arbitrariness feature) would exclude from its extension are some of the intentional gestures of great apes. Take the 'play-tap' of chimpanzees (Tomasello 2008). One youngster approaches a second, wanting to play. The first raises his arm, as though to hit the second, and the recipient begins to play upon perceiving this initial arm raise. Note how closely the physical form of the signal is tied to its function. Tomasello (2008, p. 26) expresses the point by stating "the 'meaning' or communicative significance of [the gesture] is inherent in [it]". The (relative) non-arbitrariness of play taps is reflected in how the signal is learnt. Over repeated instances of one individual coming up to a recipient to initiate play by hitting him, the recipient learns to anticipate impending unruliness just from a raised arm. The English utterance 'let's play' is very unlike a chimp play tap in this respect. The physical form of this utterance has no intrinsic connection with playing (which is why the content '*let's play*' can be expressed in many different languages, in many different tones of voice, at many different volumes, etc.).

In response to these cases, there at least two possible ways forward. The first is to give up on an information-mediated influence definition of communication to include cases like the bee dance, indices of quality, and certain intentional gestures. The second option, however, is to 'bite the bullet', keep the definition, and instead rule out (as non-communicative) cases like the bee dance, indices of quality, and intentional great-ape gestures like the play-tap. After all, only if ethologists are using the notion of 'communication' consistently would we expect to be able to define communication in way that perfectly captures biological practice. One might be inclined to question this assumption.

I suggest the following. We need not assume from the outset that a definition of 'communication' must capture all biological applications of the concept. There can be room for theory to influence scientific practice and scientific intuition. But there are limits. It can't be the case that too many interactions classed as communitive are ruled out. Nor should we go so far as to rule out interactions considered to be paradigm cases of communication. The bee dance, for example, is one such case of a paradigm communication system in biology.

4. Conclusion

In this chapter I have argued that an information-mediated influence definition of animal communication (as presented by Scarantino 2013) is extensionally inadequate. It is too liberal, in that it includes various co-adapted interactions which are normally considered to be 'non-communicative'. I then investigated whether arbitrariness could be added to an information-mediated influence definition in order to exclude various 'non-communicative' interactions. I concluded that arbitrariness is unhelpful because it swings things too far in the opposite direction, excluding many stock-and-trade cases of communication. Of particular concern is that it excludes some paradigm cases, like the bee dance. We go from a definition of communication that is too liberal to one that is too restrictive.

It might be beneficial at this point to take a step back and assess the general strategy. An adapted influence definition of communication (MSH 1995), a co-adapted influence definition (MSH 2003, Scott-Phillips 2008), and an information-mediated influence definition (Scarantino 2013)[41] are all definitions. Yet, constructing a definition is only one way to go about the process of categorising something (like 'communication'). It reflects the 'classical' approach to categorisation (Margolis & Laurence 1999): come up with a list of necessary and sufficient conditions that (hopefully) include all and only instances of communication.

However, it is no secret that variation characterises many biological phenomena. For example, a list of necessary and sufficient conditions is unfit to

[41] Including an information-mediated influence definition that includes an arbitrariness feature.

categorise species (Hull 1986). Nor is such a list apt to capture the notion of a 'Darwinian population' (Godfrey-Smith 2009). In cases like these, classical categorisation is hard because there are many intermediary cases between 'full-blown' instances of a category, on the one hand, and partial instances of the category, on the other[42]. However, if this chapter is on the right track, then categorising animal communication is hard for a different reason. It is hard, not because there are many intermediate cases of communication lying mid-way between (say) co-adapted interactions, on the one hand, and 'full-blown' communication, on the other (although this might be an additional reason). Rather, I have sought to show that classifying communication is hard because the interactions many want to label as properly 'communicative' are, upon closer examination, simply no different from co-adapted interactions more generally.

The main moral is that neither information, (co-adapted) influence, nor a combination of the two, is a defining feature of animal communication. If I am right, then it is extremely difficult to specify what*,* exactly, makes animal communication a unique kind of interaction. It can't be because animal communication is merely a co-adapted interaction. This is too liberal. And it can't be because it is a co-adapted interaction that involves the transmission of natural meaning (where natural meaning is understood as a structuring cause of receiver behaviour). This is because, as I have argued, transmitting natural meaning is a feature of most (if not all) co-adapted interactions.

The idea that 'information transmission' or 'carrying information' is not special to co-adapted interactions that are considered communicative might strike some as initially strange, particularly when considered against the backdrop of human communication and linguistic communication. There seems, at least on prima facie reflection, to be something utterly unique about language. Language seems to be an exceptionally powerful communication system. As well as expressing feelings and emotions, it allows us to communicate about states of affairs massively displaced in both space and time, and even completely non-existence states of affairs. As a result, it is used to strengthen social bonds (gossip), do science, create poetry, and to coordinate complex behavioural interactions like hunting or building bridges. Indeed, without language the kind of cultural evolution unique to our species would probably be impossible. Surely the extraordinary power of language comes down to the fact that language conveys information in some unique sense?

Maybe so. Perhaps the way in which human language conveys information is totally unique. However, even if this is true, it does not mean that nonhuman signalling (and non-linguistic signals among humans, like facial expressions) is unique

---

[42] I thank an anonymous reviewer for pointing this out to me.

in conveying the kind of information that *it* trades in. If the argument I have presented in this chapter is correct, then the great majority of signals sent in the nonhuman realm carry a kind of information which is non-unique to communication. Indeed, it may even turn out that there is nothing uniquely special about (nonhuman) communicative interactions vs other kinds of co-adapted interactions among nonhuman animals.

Finally, I hope that my analysis will be of some interest to those sceptical of the idea that signals 'transmit information', more specifically of the idea that there is any explanatory utility in ascribing informational properties to signals. If I am right, then of course information transmission won't be a defining feature of communication. But nevertheless, it will still be the case that signals transmit information in the predictive sense proposed by Scarantino, the causal-explanatory role of which I have sought to clarify in this chapter.

Crucially, this view of the causal-explanatory role of natural meaning need not commit us to the idea that signals have the selected function of transmitting information. The selected function of signals can still be to influence receivers, as opposed to inform them (Dawkins & Krebs 1978, Owren et al. 2010; Carazo & Font 2010[43]). It's just that receivers were rewarded for paying attention to signals only when signals coincided with their 'referents' either over phylogenetic or ontogenetic time. Carrying natural meaning can thus be regarded as a 'normal condition' (Millikan 2004) in explaining the phylogenetic or ontogenetic development of communicative behaviours, as opposed to the selected function of a signal.

*   *   *

Chapter 3. *What is it to explain a receiver's response by positing that it 'derived' natural meaning from a signal? What is the explanatory role(s) of natural meaning when it comes to understanding receiver responses? Might getting clearer on this go some way towards resolving the information vs influence debate?*

[43] As Carazo & Font (2010, p. 663) highlight, "the function… and information content of a signal are not the same thing, as evinced by the fact that the same effect (e.g. to intimidate a rival) can be achieved by a signal with different informative content (e.g. size, social or residential status)."

Chapter 3

Information, Influence, and the Causal-Explanatory Role of Natural Meaning in Understanding Receiver Responses

As part of the move away from group selection, Dawkins & Krebs (1978) famously argued that signalling systems are not systems of information flow. Because senders and receivers are often in competition, we should instead see signals as attempts by senders to manipulate or influence receivers. Their position was soon criticised for framing receivers as passive dupes, rather than active participants with *their own* selfish interests (Zahavi 1975, Maynard-Smith & Harper 2003). Rather than viewing receivers as passive dupes, we should expect receivers to ignore dishonest/manipulative signals, forcing communication towards honesty, and thus informativeness. According to more recent sceptics of information, however, there is no justification for this "a priori" assumption that receiver selection wins out over selection for sender influence (Owren, Rendall & Ryan 2010: henceforth 'ORR'). If receivers are successful in seeing off sender influence in one domain or modality, selection should trigger "assaults" in other domains. Instead of honesty, we should commonly expect an equilibrium point where receivers are "somewhat disadvantaged".

If this is indeed the case, then the default assumption of information proponents leads to bad biology. It diverts attention from what is most important: features of signals designed to exploit sensory and psychological biases in receivers, and to facilitate detectability, localizability and avoid receiver habituation. An example is the presence of photoreceptors in the sensory systems of some fish that evolved to detect prey. These photoreceptors secondarily produced selection for corresponding coloration in males: specifically, for males to become more visually salient to females possessing the photoreceptors. But according to ORR, this change in males is neither 'honest' nor 'dishonest'. It isn't explained by the signal correlating reliably with male quality. Instead, it is explained by the signal being linked to sensory biases in female receivers. Another example is a species of frog whose sexually selected vocalisations evolved to match pre-existing auditory sensitivities in females. What explains the physical structure of the male's signal is not information about male quality, but rather the latent sensitivity of the female's auditory system. Males exploit this sensitively for their own ends. According to ORR, the assumption that signals are designed to transmit information blinds researchers to crucial features of signal design, and to the evolutionary rationale for these features.

ORR are right to stress the importance of influence in understanding signal design. However, an informational approach is not thereby discredited. I think we should conclude that it is unwise to frame the selected function of signals *wholly* in terms of honest information transmission. But information transmission must play *some role* in the evolution of signal form, for two reasons. Firstly, and as ORR themselves admit, a signalling system in which perceivers are *strongly* exploited is unlikely to persist. Secondly, something needs to explain why a receiver with a particular sensory bias, for instance, responds *in the particular way* it does. Male fish pigmentation must be designed against the background of female sensitivity and processing, if only to ensure that females take notice of males looking to mate. Any male, even a weak one, able to produce a signal that females find easier to notice will increase his fitness. ORR are right that, at this point, a focus on information is unilluminating. However, we still need to explain why the female fish responds with sexual receptivity, as opposed to some other behaviour, upon seeing the male's signal. *This* is where information content, in the sense outlined in Chapter 1, becomes crucial. Even if there is substantial (non-informative) selection pressure on signals to influence receivers, this doesn't make full sense of the receiver's response. Why do females respond to attention-grabbing signals with sexual receptivity as opposed to predation, for example? Of course, it's possible that females assess males based on other factors after receiving the initial signal, and that, as a result, the signal need not correlate reliably (or fairly reliably) with male quality. But in the absence of this, females become sexually receptive to the signal because, when things went right for females in the population over phylogenetic time, the signal correlated with a male of decent quality looking to mate. If females are responding with sexual receptivity as opposed to predation, this must be because, *often enough,* this response paid off. And responding to the signal with sexual receptivity would only have paid off if it was produced, often enough, by a male of decent quality looking to mate. This means that males of low quality must have been excluded, often enough, from signalling.

Thus, an informational approach need not discourage a focus on the intrinsic features of signals. For it encourages searching for features of the signal that restrict access to 'dishonest' senders. If there is selection pressure on receivers for scepticism, that makes the information content of the signal relevant, even if that selection pressure is sometimes not very effective, leading to sub-optimal decision making by receivers.

ORR also critique an informational methodology on mechanistic grounds. They see an informational approach as inflating the mental sophistication of receivers.

Information encourages viewing the mechanisms underpinning responses as being unduly close to the relevant ones underpinning human language. We should instead view the former mechanisms as effect-based, as opposed to 'cognitive'. Animal signals, due to their intrinsic physical properties, evoke attention and arousal in receivers directly and without cognitive mediation. Again, senders have the upper hand, but this time proximately as opposed to ultimately. In contrast, information proponents argue that, at least in some species, receivers control how they respond to signals in real-time (Seyfarth et al. 2010). In such species, signals with similar physical properties can lead to different responses, depending on what mentally-represented information is brought to bear on the information derived from the signal. Similarly, signals with different physical properties can lead to the same responses, again depending on what mentally-represented information is brought to bear on the information derived from the signal.

Bringing evolutionary and mechanistic considerations together, what motivates information proponents is essentially the idea of receiver flexibility. On the evolutionary front, proponents argue that putting senders in charge ignores that natural selection acts on receiver strategies as well as sender strategies. If being influenced by a signal in a certain way is not adaptive for a receiver, then natural selection will lead receivers to change their response. On the mechanistic front, information proponents argue that, at least in some species, receivers possess the cognitive sophistication to control their responses in real-time. If receivers are not prisoners of sender influence, then we must infer that receivers respond to signals, and respond in the particular way they do, because information is 'derived' from signals. Because of the view that receivers possess response flexibility, Seyfarth et al. (2010) state "the concept of information plays a central role in studies of animal communication".

I largely agree with information proponents. However, they need to make some important distinctions. The response flexibility of receivers can take one of two forms, corresponding to two different ways information is 'derived' from signals. As already stated, information proponents argue that there is good reason to think that evolution works on receiver strategies at the population-level to minimise malign signal influence. I will label this kind of receiver flexibility *diachronic* flexibility. Furthermore, information proponents argue that in some species receivers themselves (i.e. at the individual-level and in real time) possess the cognitive sophistication to respond to flexibly to signals. I will label this *synchronic* flexibility. The main goal of this chapter is to show that information proponents do not differentiate clearly enough between these two different kinds of flexibility; and, concomitantly, between two different ways information may be said to be derived from signals.

As I will show in the final section of this chapter, it can be unclear which of the following things information proponents mean when they say that a receiver has 'derived information' from a signal: First, that natural selection, over phylogenetic time and at the population-level, has made use of a correlational link between signal and relevant world-state in order to install or maintain (in the receiver) a relatively fixed response. If the environment has not changed drastically, then this response will presently be adaptive. The second thing information proponents might mean by saying that a receiver has 'derived information' from a signal is this: that receivers (on the individual-level) represent a correlational link between signal and world-state, in such a way that this information is made available for real-time decision-making. The second option brings with it more substantive proximate-cognitive commitments which I will identify using Sterelny's (2003) notion of *decoupled mental representation*. When no such proximate-cognitive commitments are intended, and when it is not made explicit *in what sense* a receiver has 'derived information' from a signal, confusion can result. It can sound like animal communication is being anthropomorphised, which is one of the complaints made against attributing content to signals in the first place (ORR, Rendall & Owren 2013).

My main aim in this chapter, then, is addressed to information proponents. However, a subsidiary aim of this chapter is addressed to the sceptics. My subsidiary aim is to make the point that information can be attributed to signals without depicting the proximate mechanisms of animal communication as being unduly sophisticated or otherwise similar to those unpinning human communication. When it is done as part of an ultimate explanation of diachronic flexibility, attributing information to a signal does not entail that senders and/or receivers are cognitively sophisticated. Nor, as I will argue, does doing this entail that receivers are on par with senders, when it comes to evolutionary considerations. Thus, I am critical of both information proponents as well as sceptics. However, I side with proponents of informational terminology. Information (i.e. natural meaning) is indeed useful in explaining why receivers respond, given receiver flexibility. However, receiver flexibility can be explained in ultimate terms or in proximate terms, corresponding to two different ways in which receivers use this natural meaning.

My plan is as follows. Section 1 explains how natural meaning can be evoked as part of an *ultimate* explanation of *diachronic* receiver flexibility. Section 2 explains how natural meaning can be evoked as part of a *proximate* explanation of *synchronic* receiver flexibility. Finally, Section 3 addresses a central pro-information paper in the debate where the distinctions I highlight are not clear: that is, where it isn't clear whether natural meaning is being evoked as part of a proximate explanation of synchronic flexibility, or merely as part of an ultimate explanation of diachronic flexibility.

First, I will explore how natural meaning is often attributed to signals as part of an ultimate explanation of diachronic response flexibility. Here, this meaning is represented by the receiver in a minimal sense only. It is natural selection on the population-level, as opposed to the individual organism, making use of the natural meaning transmitted by signals.

1. Natural Meaning in Ultimate Explanations of Diachronic Flexibility

Anolis cristatellus lizards are preyed upon by a species of snake, *Alsophis portoricensis*. When it spots a snake, a lizard will perform a conspicuous display: push-ups, in which it moves its body up and down in a vertical plane by flexing and extending its legs. *A. cristatellus* lizards escape from snakes in one of two ways. Sometimes, they flee at the moment of attack. More often, however, they must struggle and bite their attackers once caught. Such struggles are usually violent, involving a significant aerobic component. Thus, push-up displays are performed at the very moment a lizard needs all the energy it can muster. Why, then, are they performed?

One explanation for this expensive behaviour is that lizard push-ups are 'honest' (or fairly honest) signals of escape-ability, which communicate to snakes that a struggle is not worth the effort. Indeed, during predation, snakes respond to lizard push-ups by stopping their approach. Additionally, lizards that signal are attacked significantly less often than lizards that don't signal (Leal 1999).

If this honest signalling hypothesis is right, then a relationship should exist between the intensity of a push-up display, on the one hand, and the ability of the signalling lizard to escape attack, on the other. As it happens, the endurance capacity of *A. cristatellus* has been investigated in the lab. The intensity of signals, represented by the number of push-ups given by a lizard during predation episodes, has been found to correlate significantly with individual physiological condition, represented by endurance (Leal 1999)[44]. Thus, there is reason to think that those push-up displays that cause a snake to disengage carry information about the sender being physiologically robust and likely to escape.

1.1 The content of a push-up display

A push-up display can naturally mean the sender has a high escape ability without entailing high escape-ability. The probability of high escape-ability, given the display, just needs to be greater than the probability of high escape-ability in the absence of the display. On this view, if a push-up display raises the probability of high-escape

[44] The ingenious study tested endurance capacity by encouraging lizards to run around specially engineered racetracks.

ability from (say) 50% to 80%, the display carries information about high escape-ability. In addition to this correlational information on the input side of equation, we can focus on the output conditions of the signal: i.e. the selected function of the receiver in response to signals (Millikan 2004; Stegmann 2009). Out of all the states of the world a push-up display correlates with, including the presence of a snake, the content of a signal is identified with that state which rationalises the receiver's response. What state of the world made it such that a snake's response to a vigorous push-up display was adaptive over phylogenetic time? Not the presence of a snake. Rather, the signalling lizard having a high endurance capacity. This, then, is the natural meaning of the signal according to the content-recipe defended in Chapter 1.

1.2 CONTENT AS USED BY NATURAL SELECTION

What explanatory work does the notion of content do in explaining the response behaviours of snakes in our example? Upon observing a vigorous push-up display, for example, it is intuitive to say that a snake disengages *because* the push-up display 'transmitted' or 'carried' information about the high escape-ability of the lizard. But what facts about the world ground this sort of 'because'?

A push-up display is a phenotypic trait of a lizard, the function of which is to cause predatory snakes to disengage. Snakes, in turn, possess the biological function of responding to vigorous push-up displays by disengaging. We can plausibly assume that the mechanisms underlying response behaviour, shaped by natural selection, are wired quite directly to certain behaviours or bodily process, and that such mechanisms are relatively automatic and ballistic. For instance, there is no indication that a snake takes into account a significant range of information, in addition to a push-up display, before assessing in a top-down manner whether to disengage from a hunt. Concomitantly, there is no indication that a snake uses the information gained from observing a push-up display to control responses other than hunt disengagement. In this sense the information carried by a signal is represented in a minimal sense only (Sterelny 2003). The correlation between signal and world-state explains why the receiver's response mechanisms are hard-wired as they are, but this information is probably not something that is decoupled from any particular response. That is, this information is probably not something the organism itself thinks about.

Instead, information comes into the picture when explaining why the signalling system is set up (and/or maintained) as it is. For instance, in explaining *why* a particular snake responds to a vigorous push-up display by disengaging, we must note the correlational link between a) vigorous push-up displays and b) high escape-ability in signalling lizards across past episodes of communication at the population-level. Historically speaking, when lizard push-up displays caused ancestral snakes to disengage from a hunt, a condition that accompanied this response, more often than not, was that the signalling lizard was physiologically robust with high escape-ability. Because of such historical and population-level facts about the system, we now say things like 'a vigorous push-up display produced by a lizard indicates that it is physiologically robust'.

It should be noted that this claim is not an entirely novel one. Teleosemanticists, when they have addressed the explanatory role of information, have noted that this role is ultimate in nature (e.g. Shea 2007; Millikan 2007). This should come as no surprise, seeing as teleosemantic theories of content all appeal to the notion of selected function. However, when it comes to purely correlational/non-etiological accounts of information (e.g. Scarantino 2015; Skyrms 2010), there has been less, if any, precise discussion of the causal-explanatory role of appealing to the information carried by signals.

In exploring this issue further, consider a relatively simple, 2 state, 2 signal, 2 act sender-receiver model involving one sender and one receiver at time T1. The sender (S) produces a signal depending on whether state 1 or state 2 obtains in the world. As a result of being influenced by a signal, the receiver (R) can act in 1 of 2 ways. Assume, further, that the system is in equilibrium such that S sends signal 1 when and only when the world is in state 1 and signal 2 when and only when the world is in state 2; and concomitantly that R performs act 1 in response to signal 1 (and only signals 1) and act 2 in response to signal 2 (and only signal 2). Finally, assume that the likelihood of either state 1 or state 2 obtaining in the world is 0.5. In this signalling system, signal 1 thus raises the probability that state 1 obtains from 0.5 to 1; and signal 2 thus raises the probably that state 2 obtains from 0.5 to 1. As a result, we can say signal 1 carries the correlational information that state 1 obtains; and signal 2 carries the correlational information that state 2 obtains (Skyrms 2010).

The question is, when R is causally influenced by a token of signal type 1 at time T1, and as a result produces act 1 at time T1, what kind of explanation are we giving of R's response at time T1 by saying '*R received the correlational information that state 1 obtains*'? Let us call the particular signal token that causally influences R, at time T1, 'Signal 10001'. The first thing to note is that the information carried by Signal 10001 is grounded by facts about the type of signal 10001 is a token of. More specifically, it is grounded by the (population-level) fact that this type of signal was given only when state 1 obtained, but not when state 2 obtained. Secondly, by invoking the information carried by Signal 10001 in order to explain R's response, we are not appealing to features of the present causal interaction between Signal 10001 and R. Instead, we are appealing to past causal interactions: more specifically between past tokens of signal type 1 and ancestors of R. We are appealing to the (historical) fact that, when R's ancestors responded to tokens of signal type 1 by producing act 1, world-state 1 always obtained.

Conversely, imagine that sender S gives, for the very first time, a token of signal type 1 in state 2 (say, accidently). In this case, the signal will carry the ('false') information that state 1 obtains, when in fact state 2 obtains. As a result, R responds by producing act 1 in state 2. What kind of explanation are we now giving of this behavioural response, by invoking the fact that R was given the false correlational information that state 1 obtains? We are basically saying that the signal hasn't co-occurred with the state that a) *past* tokens of b) this signal *type* co-occurred with. In other words, we are appealing to a) historical facts and b) population-level facts. Compare this with the way in which we explain someone's death by saying that his heart 'malfunctioned' (as opposed to giving a mechanistic explanation of what caused his bodily-functions to cease). Here we are saying something about the malfunctioning heart, but what we are saying is grounded by a) historical facts about past tokens of this organ in b) an ancestral population of organisms. Our present malfunctioning heart isn't doing what it is supposed to do. What it is supposed to do is what the numerous hearts (population-level) did in the ancestors (historical) of our unfortunate subject, such that he ended up with a heart, albeit one that just stopped working. In giving the 'false information' explanation for R's response behaviour, and in giving the 'the malfunctioning heart' explanation for our unfortunate subject's death, we are in both cases giving ultimate explanations.

1.3 Implications

The first moral is directed towards information proponents. Information is often said to be 'derived' by receivers from signals sent and received among fairly cognitively-simple organisms: organisms whose signalling strategies are hard-wired, and who map stimuli onto response behaviours in a relatively direct and automatic manner. When information is said to be derived from a signal by a receiver in such a case, this should be seen as part of an ultimate explanation of the receiver's response. The information derived from the signal is not freely available to drive decision-making in real-time. Instead, it explains why the receiver's typical response strategy takes the form that it does. Using Dretske's terminology introduced in Chapter 2, the information explains (2^{nd}-order) the conditions responsible for the current situation in which signal tokens of the relevant type cause (1^{st}-order) the typical response.

Moreover, in such cases information is central to explaining why receivers are not necessarily prisoners of sender manipulation. As shown by Leal (1999), a push-up display is a relatively honest signal of escape-ability. In other words, vigorous push-up displays correlate with high escape ability. This is almost certainly due to population-level variation on the receiver's side, allowing receivers to ignore vigorous push-up displays that do not correlate with high escape-ability. This, in turn, creates selection pressure for costly signals which ensures signal honesty (Maynard-Smith & Harper 2003; Searcy & Nowicki 2005). But note that the flexibility creating honest and informative signals exists over phylogenetic timescales, in this case. It is explained ultimately as opposed to proximately.

The second moral is directed towards information sceptics. In communication systems like the one discussed, attributing information to signals *in no way* anthropomorphises the communication system. It does not depict sender and/or receiver strategies as being driven by anything like the psychological abilities underpinning intentional human communication. Moreover, attributing information to signals does not entail that receivers are on a complete par with senders, when it comes to game-theoretic considerations. A lizard's push-up display can carry the information that it is likely to escape even if, sometimes, the lizard would not escape if pursued. Indeed, when it comes to lizard push-ups, signal honesty is enforced by the target receiver, as opposed to being ensured by cost paid up-front. As a result, we should expect some amount of dishonest 'bluffing' to occur (Bruner at al. 2017). Because a snake's evolved, species-typical, response to a vigorous push-up display was adaptive in ancestral environments when and only when the lizard was physiologically robust, the display carries information about the lizard being physiologically robust as outlined in Chapter 1. This remains the case even if relatively weak lizards sometimes 'bluff', producing a vigorous push-up display and dangerously expending valuable energy.

I will now explore how information is sometimes evoked as part of a *proximate* explanation of *synchronic* response flexibility. Here, information is represented by receivers in a more substantive sense. It factors into the real-time decision-making process of individual receivers.

2. Natural Meaning in Proximate Explanations of Synchronic Flexibility

A famous paper published in 1980 by Seyfarth, Cheney and Marler brought to the wider attention of biologists, psychologists, neuroscientists, philosophers and linguists the now infamous alarm call system of the vervet monkey. Along with this pioneering work, research on other primate species (Pereira & Macedonia 1991) and chickens (Evans et al. 1993) eventually gave rise to the idea of 'functional reference' in animal communication theory. Combined with Donald Griffin's work on using communication as a 'window' into what animals think, it also contributed to the rise of cognitive ethology (Manser 2013).

Seyfarth et al. (1980) found that vervets have distinct acoustic calls for leopards, eagles and snakes which elicit specific and context-appropriate responses. Upon receiving a leopard call, conspecifics run up trees. In response to eagle calls they run into a bush, or if they are already in the top of a tree, they move down closer to the centre of the tree. In response to a snake call, receivers stand up bipedally and look around. These predator-avoidance behaviours are also triggered when individuals encounter the predators themselves, as opposed to alarm calls designating them. However, Seyfarth et al. managed to trigger each kind of avoidance behaviour in the absence of the eliciting predators, by playing back pre-recorded alarm calls from speakers strategically hidden in trees and bushes.

This work caused much excitement. It was taken to constitute evidence, for the first time, of 'symbolic'/'arbitrary' reference in a nonhuman species (Hauser 1996). The idea that nonhuman signals could bear a symbolic or arbitrary association with their 'referents' went against the then prevailing understanding of animal communication. The then prevailing view was that signals merely reflect the internal motivational state, or the immanent behaviour, of the sender (e.g. Dawkins & Krebs 1978; Owren & Rendall 1997). According to this non-symbolic view, signal form was seen as highly constrained. This stands in contrast to words.

A dog baring its teeth is a non-arbitrary signal of an emotional state liable to produce aggression. This is because the physical form of the signal vehicle is highly counterfactually constrained. Only a very small number of alternative candidate signal vehicles might have been used to transmit the same information, at the time of initial signal development. This contrasts with some linguistic signals, in which signal form is counterfactually constrained to a lesser degree. The physical form of the printed word 'd-o-g', for instance, could have taken a number of different forms at the time of initial dubbing (for instance 'd-a-g'). It is because of this that we have the intuition that the physical sign 'd-o-'g' carries the information that it does in virtue of *coordination* among senders and receivers, as opposed to 'physical necessity'.

Vervet alarms were found to be relatively arbitrary in this sense - in a way that, say, a threat signal from a dog is not. When it comes to the former, coordination plays a substantial role in explaining why a leopard alarm call (individuated here purely in virtue of its physical form) is now used to signal the presence of a leopard[45]. On the other hand, coordination between senders and receivers plays a less substantial role in explaining why a teeth-baring display among dogs is now used to signal the sender's aggressive emotional state.

More recently, ethologists working on the informational content of 'referential' signals adopted the watered-down phrase 'functionally referential' to refer

[45] Although here coordination is not the result of rational processes, as opposed to the law of effect.

to signals like vervet alarms (Hauser 1996). The watered-down claim that such signals possess functional reference, as opposed to full-blown reference, reflected caution relating to the mechanisms underlying signal-production. The underlying mechanisms are most likely different[46].

Signals considered to be functionally referential, like vervet alarm calls, are usually thought to display the following two features[47]. Firstly, signal production by senders must be context-specific. That is, signals must be produced in a narrow range of circumstances. Secondly, the appropriate response to a signal carried out by a receiver (say, running for cover from a predator) must occur even in the absence of the circumstances that normally cause signal production (i.e. a predator). That is, the call must be sufficient to elicit behaviour appropriate in the state of the world 'referred' to by the call (Manser 2013).

The first condition, context-specificity, arguably presents further trouble for the idea that functionally-referential systems are akin to reference in human language. As Wheeler & Fischer (2012) note, the more context-specific a signal is, that is, the more tightly its production is linked with specific states of the world, the *less* cognitive sophistication is required to respond to it adaptively[48]. Context-specificity implies that the amount of cognitive sophistication needed for a receiver to respond as though a signal refers to something in the world is not very high. A receiver need only be stimulus-bound; automatically responding to a particular signal with an invariant behaviour. This relatively simple stimulus-response connection could be formed by either evolution or simple behavioural conditioning in a learning period[49].

A similar moral can be drawn from Sterelny's (2003) work on the evolution of cognition. If cues/signals reliably correlate with various states of the world, then in order for responses to such stimuli to be adaptive, natural selection need only provide receivers with modest response mechanisms that invariantly trigger some very specific and ballistic behaviour. When signals are *not* given in very narrow contexts, i.e. where the environment is 'epistemically opaque', *then* it can become more adaptive for receivers to respond in a flexible and non-ballistic manner to incoming stimuli. So, there seems to be no necessary connection between functional reference and linguistic-like psychological sophistication (on the part of receivers, in this case). If

[46] Although Wheeler & Fischer (2012, pp. 196-197) note "the idea that functional referential signals require greater cognitive complexity and provide a clearer link to human language than do other types of animal signals remains pervasive in the animal communication literature".
[47] Although see Scarantino (2013) for an alternative proposal.
[48] And also, the less like a word it is. As Sterelny (2016), following Deacon (1997) has stressed, very few tokens of "tiger" are uttered in the presence of tigers.
[49] There is an important caveat. Even if the receiver's cognitive mechanism for learning the signal-world correlation is more sophisticated it would still not follow that the signal would be anything like a word. If, for example, the receiver learns what the signal correlates with via a form of social learning, such as by observing the responses of other animals, this would not make the signal word-like. This is because what is being learned, however cleverly or not, is just a very specific signal-world association. And, as Deacon (1997) pointed out, word meaning is more complicated than a large list of signal-world associations.

anything, the inference seems to point the other way; towards psychological simplicity[50].

Despite the attention received by vervets relating to the issue of referential-like communication, what is significant for the current discussion is not strictly speaking whether or not signals correlate reliably with a limited range of world-states. Instead, what is significant about vervet communication is the response flexibility displayed by receivers (a feature which is not directly connected with the functionally-referential nature of alarm calls). This response flexibility was revealed by different studies carried out, again, by Seyfarth, Cheney and colleagues.

Eight years after their original ground-breaking paper on vervets, Cheney & Seyfarth (1988) presented some new research suggesting that receiver vervet monkeys mentally represent the information derived from vocal signals in a fairly rich sense. This time the research was carried out on inter-group contact calls, as opposed to alarm calls. Individual vervets would learn that a particular vocal signal normally indicating contact with another group and coming from a particular conspecific was unreliable. As a result, they would become bored with the signal and soon ignore it. In other words, they habituated to the signal. However, as a result they also came to ignore an acoustically different signal given soon after by the same conspecific - so long as the two acoustically different calls carried the same information (in this case, about the presence of another group of vervets nearby). If, however, the two different acoustic calls carried different information, then habitation was not transferred. Moreover, if the identity of the sender changed between the first and second call, habituation was not transferred.

The idea is as follows: if habituation to a first signal produces habituation to a second, receivers have judged the two signals as being the same in some way. There are two different ways in which two signals might be judged the same in the mind of a receiver. First, they might be grouped based on their similar acoustic features. Second, they might be grouped based on their shared informational content (if, in fact, they share content). If the two signals sharing content are grouped together while at the same time differing markedly in acoustic features, then they must have been grouped together because of their shared content.

In Cheney & Seyfarth (1988), an initial inter-group contact call type C1 was produced as though coming from sender S1, but in the absence of inter-group contact (its normal content). It thus constituted ‘fake news’ for the receiver R. As a result, R habituated to it. One deflationary hypothesis is that R was simply habituating to the proximal nature of the stimulus, as opposed to its meaning. However, if this was the explanation then R should presumably not have habituated to an acoustically different

[50] For debate about the continuing importance of the functional reference paradigm, see Wheeler & Fischer 2012 and the subsequent exchange between Scarantino & Clay 2015 and Wheeler & Fischer 2015.

call, C2, which nevertheless carried the same information. Instead, R should have become vigilant as though another group was nearby. Instead, R remained bored by this new signal, as though R knew what distal world-state it was supposed to indicate (along with the prior knowledge that this world-state was not actual). Moreover, as noted, if instead the second signal-form C2 was produced as though coming from a different sender S2 such that the identity of the sender changed between C1 and C2, R took tokens of C2 seriously as though another group was indeed nearby – despite not taking C2 tokens seriously if produced by S1. In other words, despite having come to distrust the information from S1 that another group was incoming, R suddenly took this information seriously when it appeared to come from a different sender. According to Cheney & Seyfarth, this suggests that R was mentally representing both the reliability of the sender as well as the distal world-state normally indicated by the signal.

The same kind of study was carried out a decade later by Zuberbuhler, Cheney & Seyfarth (1999), this time on Diana monkeys. The study provided evidence for the same kind of rich processing of signals on the part of receivers. Here, Diana receivers were played sequences of 2 vocalisations separated by 5 minutes of silence. There were four different vocalisations to draw from to create a sequence of 2 vocalisations. The four possibilities were, firstly, the growl of a leopard; second, the shriek of an eagle; third, a conspecific Diana monkey's leopard alarm call; fourth, a conspecific Diana's eagle alarm call.

In the baseline condition, receivers heard a predator call from a conspecific followed by another token of the same predator call from the same sender. Because the acoustic and the semantic features remained the same across the 2 vocalisations, habituation was transferred from the first vocalisation to the second. In the test condition, however, receivers heard a conspecific Diana's alarm call, this time followed by the corresponding vocalisation of the actual predator. Here, only the semantic features remained the same across the 2 vocalisations, while acoustic features differed. Interestingly, it was found that in this case habituation was transferred from an alarm call to the vocalisation of the actual predator – despite the different acoustic properties of the stimuli. According to the authors, this suggests that Diana monkeys mentally represent the world-state indicated by signals in a fairly rich manner.

Here we have two examples of communication in which receivers display a significant amount of real-time flexibility in their response profiles. The moral, to be fleshed-out now, is that in cases like these, information is derived by receivers from signal-world correlations in a very different sense to the way information is derived by receivers from lizard push-up displays.

2.1 Mentally representing the content of signals in a rich sense

In the last study discussed above, the authors conclude as follows:

> When hearing natural stimuli, monkeys do not rely only on the acoustic features of these stimuli alone to govern their behavior but instead access a common associate, possibly a mental representation, of the predator category. Perhaps these mental representations are not unlike those linked to the human linguistic concepts of *leopard* and *eagle*. Zuberbuhler et al. (1999).

There are many ways of understanding the notion of mental representation. In cognitive science, one traditional way is to view mental representation as an internal 'language of thought' (Fodor 1975). According to this view of the mind, mental states like beliefs and desires are structured symbolic representations composed out of more basic symbolic states called *concepts*, analogous to how sentences are systematically composed out of more basic representational units (i.e. words).

In comparative psychology, however, the received view is somewhat different. The notion of a concept/representation is operationalised. Organisms possessing mental representations behave flexibly, as opposed to rigidly. Their behaviour goes beyond a one-to-one mapping between stimuli and response. E.O. Wilson (1971) presented the case of the *Pogonomyrmex barbatus* ant's reaction to dead conspecifics. Workers respond to the dead body of one of their sister workers by grasping it and carrying it away to a waste pile. Dead conspecifics are identified via a chemical product that they produce. However, workers are relatively unsophisticated in their identification abilities: almost any object tainted with a similar chemical product will be treated like a corpse and relegated to a waste pile. Thus, worker ants seem to lack anything like a 'concept' of a dead conspecific. They respond invariantly to a single cue that is normally co-extensive with dead conspecifics, regardless of whether the cue actually co-occurs with a dead conspecific at that time.

On the present view, an organism that possesses a genuine concept - say, of a kind of predator - should respond in a functionally relevant manner to that kind of predator across a range of perceptual stimuli. In other words, its behavioural profile must be at least many-to-one: adaptively mapping various kinds of perceptual stimuli onto a particular response. However, this many-to-one behavioural mapping must go beyond *perceptual stimulus generalisation*. Perceptual stimulus generalisation involves an organism extending control over some response from a class of stimuli previously associated with the response to some novel stimulus perceptually like the previously-associated stimuli. The generalisation is made possible because the novel stimuli is perceptually like the previously-associated stimuli.

For instance, if a dog has been conditioned to fear the sight of its cruel owner who happens to have a large black beard, it may also display a fear response to a fake black beard attached to a mask worn by someone else. The behaviour of parrots in categorisation tasks in the lab indicates that they engage merely in perceptual stimulus generalisation. After being trained to group pictures of different kinds of trees under the one category, they then overgeneralise and include broccoli due to the crude perceptual similarity between a picture of a tree and picture of a broccoli.

The acid test (Shettleworth 2010) of concept possession in comparative psychology is the ability to generalise, not merely to novel stimuli that are perceptually simular to a stimulus previously associated with a response, but rather to novel stimuli that share an abstract or higher-order similarity with a previously-associated stimulus (see also Allen & Hauser 1991). Humans, for instance, place broccoli in a different category to trees, due to functional and other higher-order differences between trees and broccoli.

In addition to higher-order, many-to-one, categorisation abilities on the input side, Sterelny (2003) and Bermudez (2003) have argued that concept/representation-possession is evinced by a behavioural profile characterised by significant flexibility on the output side of behaviour, as well as the input side. Sterelny (2003) identifies a series of incrementally more sophisticated ways in which an organism might represent its environment. The baseline is an organism which conditions a relatively fixed response on only one kind of cue. The next step up is an organism which engages in 'robust tracking'. Robust tracking involves an organism conditioning a fixed response on more than one sort of cue (i.e. many-to-one categorisation). Conditioning behaviour on multiple cues allows for informational redundancy in 'opaque' environments: i.e. environments in which information from any single cue might be unreliable. This might be due to environmental fluctuations or deception from hostile organisms. However, response breadth might vary independently of input redundancy. Thus, a creature might face an opaque environment but still perform a relatively fixed behaviour when a world-state has been identified as obtaining based on multiple cues. Alternatively, some organism might respond flexibly to a range of incoming cues *with a range of behaviours* (see figure 1). This is the next incremental step above robust tracking. Sterelny calls it 'decoupled mental representation'. In a creature possessing decoupled mental representations, the environment is represented both a) robustly and b) in such a way that these representations are not tied to any one response.

When it comes to the communicative behaviours of receiver monkeys, Cheney & Seyfarth (1988) and Zuberbuhler, Cheney & Seyfarth (1999) seem to highlight a many-to-many response profile indicative of decoupled mental representation. For instance, a Diana receiver can use a range of quite different stimuli (i.e. eagle alarm calls or the sound of an actual eagle) to trigger a particular response (i.e. running into a bush). Also, it can respond in a range of different ways (i.e. looking up or running into a bush) to a particular kind of stimuli (i.e. an eagle alarm call).

Applying Sterelny's (2003) terminology, we can thus say that a Diana monkey has a decoupled mental representation of an eagle.

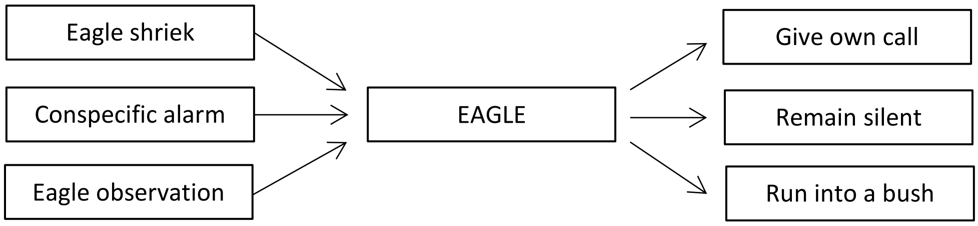

Fig. 3.1 A many-to-many behavioural profile indicative of decoupled mental representations

2.2 INFORMATION AS USED BY THE INDIVIDUAL ORGANISM

Information proponents studying animal communication invoke decoupled mental representations. They do this primarily in order to explain the flexibility inherent in the comprehension side of nonhuman primate vocal communication. The comprehension of vocal communication in primates is plastic over ontogeny, with receivers having to learn what signals designate. This stands in contrast to the production side, where the contexts that elicit signal production usually come hardwired from birth and do not change much over ontogeny[51]. That is, receivers learn their signalling strategies to a significantly greater extent than senders. This asymmetry in vocal communication is widely noted in the literature (e.g. Seyfarth & Cheney 2003, 2010; Zuberbuhler 2005; Tomasello 2008; Fitch & Zuberbuhler 2013)[52].

In what initially seems puzzling, Seyfarth & Cheney (2003, p. 153) argue that the consumption side of vervet communication is representational because the production of alarm calls is "strongly associated with a very narrow range of eliciting stimuli". However, we saw previously that whenever signal production is bound tightly to particular contexts, we don't need to appeal to cognitive complexity in order to explain adaptive responses (Sterelny 2003; Wheeler & Fischer 2012). Seyfarth & Cheney (2003, 2010), however, endorse a cognitively rich account of Pavlovian conditioning. For them, learning to associate a signal with some state of the world involves an organism conjoining two mental representations of its environment: a

[51] Vervet infants initially give alarm calls to warthogs and pigeons that pose no danger to them. But such mistakes are not random: infants 'over-extend' leopard alarm calls to terrestrial mammals and eagle alarms to birds. From birth then, vervets divide predators from non-predators, and within the predator category they divide terrestrial carnivores from aerial predators (Seyfarth & Cheney 2010).

[52] This is not to say that signal production in old-world monkeys is inflexible and insensitive to context. Senders do not have to learn the circumstances under which a signal is to be produced, but nevertheless they often must take into account contextual features of the situation to determine whether or not to signal*, at any particular time*. Audience effects are a good example: senders are sensitive to conspecifics, and adjust when they produce signals based on their presence or absence (Seyfarth & Cheney 2010, Fitch & Zuberbuhler 2013).

representation of the signal, on the one hand, and a representation of the indicated world-state, on the other[53].

The idea that Pavlovian (or classical) conditioning involves an organism 'representing' its environment has become the received view in learning theory (Rescorla 1988; Gallistel 2003; Holland 2008). The idea that Pavlovian condition is simply the registration of co-occurrence is an outdated view of the process of conditioning. The alternative, representational, view is based on findings that organisms don't associate any two stimuli that simply co-occur in experience. On the representational view, organisms come to associate a novel stimulus with some other event only if the novel stimulus provides the organism with information about the event, in a sense strikingly congruent with Claude Shannon's (1948) notion of uncertainty reduction. That is, what is required for an organism to learn the connection between two stimuli is contingency (where a stimulus alters the probability of the other event), as opposed to contiguity (where a stimulus simply co-occurs with the other event).

Rats, for example don't associate a novel stimulus (say, a tone) with a well-known stimulus (say, a mild shock) if the shock occurs frequently in the absence of the tone. This is true even if the tone, whenever it is presented, always co-occurs with the shock. Rather, the tone must raise the probability of the shock for the two stimuli to be associated by the rat. Moreover, if a novel stimulus, S2, is presented alongside a familiar stimulus, S1, where S1 is known to be connected with some salient event, E; and if the organism is then presented with both S2 and S1 whenever E occurs, it will not come to associate S2, the novel stimulus, with E. This is because S2 provides no (new) information about E that S1 doesn't provide. That is, S2 doesn't decrease the organism's 'uncertainty' about E (because S1 already has). This phenomenon is called *blocking*. Similarly, when two novel stimuli are presented together for association with some other event so that there is redundancy, the organism ignores one of the novel stimuli because it provides no (unique) information. This is called *overshadowing*. It is very similar to blocking, the only difference being that overshadowing involves the presentation at the outset of two as yet non-associated novel stimuli contingent with

[53] It was argued previously that the idea of functional reference is a red herring when it comes to the psychological sophistication of animal communication. This was because the context-specific nature of signals thought to 'refer' in fact *rules out* the likelihood that response mechanisms are sophisticated or language-like. The view being considered now is that vervet communication, for example, is of interest because signals correlate reliably with distal states. This seems like a contradiction. However, note that the mechanisms which are sensitive to the tight signal-world relation in the case of vervets are ontogenetic (on the consumption side of communication), while the functional reference paradigm would characterise hard-wired response mechanisms as possessing some interesting kind of sophistication. The implication is that if Pavlovian conditioning in some animals is representational, as will soon be discussed, then functional reference might be significant for the issue of human-nonhuman communicative continuity if it restricts itself to context-specific signals that are *learnt*, as opposed to context-specific signals that are developmentally canalized.

some further event, whereas in this case of blocking one of the two stimuli is already associated with the event at the outset.

Other findings related to how animals associate stimuli have led to the now widespread view that Pavlovian condition is a representationally-rich process: indeed, one involving decoupled mental representations underpinning many-to-many response profiles. Assume that some organism has come to associated S1 (say, a tone) with S2 (say, a shock), based on the tone raising the probability of a shock. That is, the probability of the shock, given the tone, is higher than the probability of the shock alone. Also, suppose the tone is not made redundant or blocked by some other stimuli. After a learning period in which the organism has associated the tone with the shock, suppose the organism now responds to the tone with fear. There are at least two possible reasons why the organism might respond with fear to the tone. First, because the organism's pre-existing response to the shock (i.e. fear) has simply been connected to the tone during the learning episode. Second, because the organism has connected separate mental representations of the tone, on the hand, and the shock, on the other.

The difference here is important. It is between a picture of Pavlovian conditioning in which the internal association forged during learning is 'stimulus-response' (S-R), versus one in which the internal association forged is 'stimulus-stimulus' (S-S). S-S forms of learning are representationally rich, while S-R forms are not (Holland 2008). S-S learning involves the organism associating a mental 'image' or 're-presentation' of the unconditioned stimulus (the tone) with an image or re-presentation of the conditioned stimulus (the shock). S-R learning involves the organism directly wiring an old response (fear) to registration of the unconditioned stimulus (the tone). S-S learning involves greater psychological mediation, in that the response elicited by the novel stimulus (the tone) is a product of the novel stimulus activating a mental representation of the conditioned stimulus (the shock).

The issue concerns the structure of the process by which Pavlovian conditioning effects behaviour. Does the tone simply become capable of evoking the response originally under the control of the shock? Or does the tone become associated, not merely with a pre-existing response to the shock, but rather with a representation of the shock? If the latter option is actual, then the organism can respond to the tone such that the tone *predicts* the shock, as opposed to merely responding to the tone in the same way it would upon getting shocked.

The latter, S-S, option is more like the kind of many-to-many response profile depicted in figure 1. We can think of the contrast between S-S learning and S-R learning as follows: After I learn that the weather report's forecast of rain is predictive of rain, because a rain forecast raises the probability of rain, I don't respond to reports of rain by immediately putting up my umbrella in my lounge room. Instead, upon receiving a rain forecast, I respond by planning for rain. I put my umbrella with my bag so that I take it with me when I leave in the morning, or if I don't have an umbrella, I get my coat out and hang it on my door.

Interestingly, it has been found that after an organism has associated a tone with a shock, where the natural response to a shock is abruptly increased activity (no doubt to get away), it responds to the tone not with increased activity but with dramatically reduced activity (Rescorla 1988). It has also been found that two different stimuli predictive of some salient event can evoke quite different responses. For example, rats that have associated both a tone and a localized prod with a shock respond to the tone with immobility, but respond to the prod with attempts to hide the prod from view by using any available material they can find to cover it up (Rescorla 1988). These kinds of findings show that Pavlovian conditioning is S-S: involving an organism associating a novel stimulus with a (decoupled) mental representation of another stimulus, as opposed to simply associating a novel stimulus with the response normally elicited by the unconditioned (i.e. familiar) stimulus.

The point of this detour into learning theory was to illustrate the kind of proximate mechanisms taken by information proponents to underlie the flexible responses of some primate receivers during communication. Cheney & Seyfarth ground their notion of mental representation using an S-S theory of conditioning. During the period in which a Diana receiver R learns the connection between conspecific eagle alarm calls and the presence of actual eagles, R does not simply wire the novel stimulus (i.e. eagle alarms) with a pre-existing response (i.e. eagle-avoidance behaviour). Instead, R wires the novel stimulus with an EAGLE mental representation which is decoupled in Sterelny's (2003) sense. Thus, when R hears an eagle alarm subsequent to learning, R can display 'foresight'. R can give its own eagle alarm call if the presence of an eagle is new news. Alternatively, R can remain silent if the presence of an eagle is old news.

The important point is that R's behavioural response is not determined by any particular input R receives. This is because learning how to avoid eagles in ontogeny involves, for a vervet, an underlying process more sophisticated than the wiring of a novel input to a particular response. The link between input and response, formed during ontogeny, seems instead to be mediated by a decoupled representational state. The result is a many-to-many mapping between inputs and outputs in the sense depicted by figure 1.

2.3 IMPLICATIONS

The first moral to be drawn from this section is directed towards information proponents. Information is sometimes said to be derived by receivers from signals in a cognitively-rich sense. At least in old-world monkeys, adapting to information in ontogeny as opposed to phylogeny is a cognitively-mediated process. That is, it results in the organism flexibly representing the contingent relationships in its environment. This rich way of representing the information derived from signals contrasts with the way in which receivers derive information from the likes of lizard push-ups.

Thus, the information carried by signals can be used in different ways by different organisms, depending on how they adapt to this information and whether this information is represented in a rich sense or only minimally. In species where strategies are hard-wired and where there are no decoupled representations, the natural meaning of a signal is represented in only a minimal sense by the organism. It is a population of organisms, as opposed to any particular organism, which has picked up on the relationship between signal and world-state. What explains the (diachronic) flexibility enjoyed by receivers in such populations is natural selection, as opposed to cognitive prowess on the individual-level. Alternatively, in at least some species whose signalling strategies are acquired in ontogeny such as old-world monkeys, the natural meaning of a signal can be represented in a rich sense by the organism. If the organism has the right kind of cognitive sophistication, then this information can be mentally represented such that it combines with other internally represented information, enabling the organism to respond flexibly and intelligently to its environment in real time.

The second moral to be drawn from this section is directed towards information sceptics. Attributing decoupled mental representations in order to explain synchronic receiver flexibility might present a genuine challenge to the idea that the proximate mechanisms underpinning signal-interpretation in monkeys are discontinuous with those underpinning interpretation in humans. If relatively sophisticated and decoupled brain states standing midway between sensory registration and behaviour must be invoked in order to explain flexible response behaviours in real time, this might indeed bring communication closer towards the realm of human communication – specifically when it comes to the issue of meaning. It is not my intention here to resolve this complex issue related to the evolution of communication. I will, however, make a few brief remarks.

On the one hand, one might think that even decoupled mental representation need not imply continuity with human communication when it comes to the issue of meaning. According to an influential line of thought on the evolution of human communication (e.g. Tomasello 2008, 2014; Scott-Phillips 2015[54]), recognising what a speaker 'non-naturally' means by uttering something requires not merely decoupled

[54] These researchers are influenced by Grice's (1957) analysis of non-natural meaning. Tomasello and especially Scott-Phillips also take inspiration from Sperber & Wilsons (1986) 'Post-Gricean' explanation of human communication which stresses the necessity of high-level recursive mindreading.

representational states *simpliciter*, but rather decoupled representations of the speaker's decoupled representational states: an ability known as 'meta-representation' or 'mindreading'. If human communication requires decoupled representations of representations, as oppose to decoupled representations *simpliciter*, then the findings presented above on the vocal communication of monkeys might be taken as interesting but relatively inconsequential for the issue of continuity. For there is not much evidence to think that monkeys are representing each other's mental states when they respond flexibly to the calls of conspecifics (Seyfarth & Cheney 2003).

On the other hand, one might reasonably think that the cognitive ability to flexibly adjust behaviour based on integrating a range of mental representations with varying contents, as vervet and Diana monkeys seem to be capable of, constitutes at least a partial step towards the 'inferential' and domain-general nature of interpretation in human communication (Fitch 2010; Seyfarth & Cheney 2017). A Diana monkey that transfers habituation from an eagle alarm call to the vocalisation of an actual eagle, despite the different acoustic properties of the two stimuli, might reasonably be thought to have engaged in a process of rational (or proto-rational) inference. On this view, mental representations are not only decoupled from any particular response, their decoupled status allows them to combine with other decoupled representations in a process of rational, or proto-rational, inference. While this inferential process might not operate over representations of others' mental states, the flexibility of this process might nevertheless represent a vital ingredient in human communication.

3. Information vs influence and receiver flexibility

It is now time to show why the distinctions made in the previous sections matter for the information vs influence debate. In a key paper defending "the central importance of information in studies of animal communication", Seyfarth et al. (2010, p. 3) state that "the concept of information plays a central role in studies of animal communication". Why? Because "both theory and data… argue against hypotheses based exclusively on manipulation or on a fixed, obligatory link between a signal's physical features and the responses it elicits" (Seyfarth et al. 2010, p. 3). In other words, a pro-information stance is justified because of the reality of receiver flexibility. Receivers display plasticity in their responses to signals: they are not necessarily subject to the whim of senders. Thus, 'information' essentially becomes short-hand for a view of communication in which sender influence is balanced by receiver flexibility. If there is no obligatory link between signal and response, then receivers are running the show, at least to some degree.

However, as I have argued in this chapter, information is applicable to two different kinds of explanation (ultimate vs proximate) of two different kinds of receiver flexibility (diachronic vs synchronic). It is important to note the ambiguity in the idea that there is no obligatory link between signals and responses. The link between a signal's physical features and the response it elicits can be flexible in two very different ways. Firstly, flexibility may or may not hold diachronically, or over phylogenetic timescales. Alternatively, it may or may not obtain synchronically, or in real-time.

In the case of diachronic flexibility, flexibility is possible due to natural selection. Natural selection can lead to a change in receivers' signalling strategies, given enough time, and assuming the current strategy is sub-optimal for receivers. As a result of this selection at the population-level, information is created. Whether this occurs is not inevitable, however. Owren, Rendall & Ryan (2010) point out that when interests diverge, we senders should be expected to tap into pre-existing sensory biases in receivers, such that receivers are prisoners of dishonest signals. The idea of receiver flexibility over phylogenetic time is, as such, a substantive empirical issue[55].

In contrast, synchronic flexibility is possible due to cognitive prowess. With the right kind of internal psychological sophistication, the link between the physical properties of signals on the one hand, and the responses elicited in receivers on the other, can be flexible in real-time. In such cases receivers possess response flexibility because the correlational link between signal and world-state, acquired in the individual's lifetime, is represented in a comparatively rich sense. The idea of synchronic receiver flexibility is a substantive empirical issue of a different sort to diachronic flexibility. It is largely a proximate issue, while the idea of diachronic receiver flexibility is largely an ultimate/evolutionary issue.

The two different ways in which receivers possess response flexibility are not always distinguished clearly by information proponents. For example, in the first section of the paper following the introduction, entitled 'information in animal communication', Seyfarth et al. (2010, p. 5) seem to defend the idea of diachronic flexibility. They state:

> Receivers are not, then, prisoners of the influence that specific acoustic properties have on their sensory systems. Instead, selection will favour receivers that act selfishly, adjusting their 'assessment rule' so that it is most effective in reducing uncertainty, or in providing them with information.

Here, the work of Grafen (1990) and Searcy & Nowicki (2005) is endorsed: work which focuses on the question of why, over phylogenetic time, signalling should be honest on average[56]. While natural selection operates on senders to influence the

[55] But so too is how much flexibility senders have to exploit these pre-existing biases in receivers, it must be noted.
[56] As noted in the introduction, information proponents take the existence of communication as evidence for signals being honest, unlike information sceptics who emphasise sensory manipulation.

behaviour of receivers to the advantage of the former (signals are, after all, phenotypes of senders), natural selection also operates on receivers so that receivers attend to signals only if it is beneficial for them to do so, on average. The normal condition for such benefits accruing to receivers is that signals correlate with appropriate states of the world. If signals are costly to produce, then it can be ensured that they do in fact correlate with the appropriate world-states. Signal costs ensure there is a correlation between signals and world-states, which can be quantified in models, and where this quantity can be labelled 'information' (of the correlational variety).

However, in the very next section, on 'receiver flexibility', the authors switch without warning to what seems to be the issue of synchronic flexibility:

> Supporting the view that selection can act just as forcefully on receivers as it does on signallers, there is ample evidence that receivers can learn to respond in specific ways to signals regardless of the signals' acoustic properties… (Seyfarth et al. 2010, p. 5).

Here, information is represented in a rich sense. The correlate of a signal is represented in a manner that is decoupled from any particular response. Confusingly, they then state the following two paragraphs later:

> Of course, there are also many cases in which acoustic features are closely linked to call function. Two examples are the relationship between call frequency and body size in frogs and toads (reviewed in Searcy & Nowicki 2005), and between formant spacing and body size in several mammals (reviewed in Fitch & Hauser 2003). In both of these cases, however, listeners attend to a crucial acoustic feature not just because it induces a 'nervous-system response' (although it may do this) but also because it reduces uncertainty (that is, provides information) about a competitor (Seyfarth et al. 2010, p. 5).

Is the second 'because' here, regarding 'uncertainty reduction', a proximate or an ultimate explanation? Presumably, the fact that signals induce a nervous-system response is the proximate reason why receivers respond. Presumably, the fact that signals reduce uncertainty is the ultimate reason why receivers like toads and frogs maintain their responses to signals. However, one could be forgiven for interpreting the second 'because' as a proximate-psychological explanation. This has the potential to make a reader think that what is being argued for is the idea that frogs and toads have decoupled mental representations as opposed to the much less controversial idea that, over phylogenetic time, low call-frequency in frogs was more often than not a feature of signals given by high quality males as opposed to low quality males.

Generally speaking, it may indeed be the case that "recipients are not powerless, unable to resist certain signals" (Seyfarth et al. 2010, p. 6). However, even if this is correct the relevant flexibility can be realised by very different means and over different timescales: diachronically, firstly; and secondly (at least in species with the required cognitive mechanisms), synchronically. Both ultimate and proximate sources of receiver flexibility are valid and important, and they may indeed overturn a purely sender-driven view of communication favoured by information sceptics. Conceptually, however, they need to be clearly distinguished. They need to be distinguished, firstly, so that they can be properly empirically investigated. Whether or not receivers in a species whose strategies are hard-wired are being exploited by hostile senders is a different issue from whether or not receivers in some other species represent, in decoupled fashion, the content of a signal (Stegmann 2013).

Secondly, they need to be distinguished in order to ward off the possibility of misunderstanding regarding the amount of cognitive sophistication attributed to animals that are said to derive information from signals. Some receivers derive information from signals in a minimal sense. Here, an invariant response is triggered by a signal. The response will occur in the appropriate world-state, assuming the environment has not changed drastically from the one in which the signalling system was stabilised. Alternatively, some receivers derive information from signals in a richer sense. Here, the natural meaning of a signal is mentally represented by the receiver in a decoupled manner. This representation combines flexibly with other decoupled representations so that appropriate action can be inferred in real-time.

*   *   *

Chapter 4. *Why attribute non-natural meaning to utterances? What is non-natural meaning and what are its psychological commitments?*

Chapter 4

A Psychological Notion of Non-Natural Meaning for Signals

> *"Infants… may on occasion, of course, learn to associate one of these noises with a perceptual event in much the same way a household pet may understand that the sound* dinner *heralds the arrival of food. But this is not language. Sounds become language for young children when and only when they understand that the adult is making that sound with the intention that they attend to something" (Tomasello 1999, pp. 100 – 101).*

This thesis is concerned with different notions of meaning and their explanatory roles. One has been a notion of natural meaning grounded by signal-world correlation and teleology. As argued in Chapters 1-3, natural meaning is important when it comes to an evolutionary understanding of communication: why do receivers respond to signals, given fitness interests often diverge? Receivers respond to signals because, when it mattered in the past[57], signals were produced in a world-state explaining why the typical response of receivers was successful. Chapter 3 then argued natural meaning is sometimes relevant to proximate explanations of why receivers in some species respond to signals flexibly, in real time. Receivers in such species learn the natural meaning carried by signals in ontogeny, where this learning involves combining decoupled mental representations. Because the correlate of a signal is represented in decoupled fashion, receivers can combine the natural meaning derived from a signal with other decoupled representations, to infer an appropriate response 'on the spot'.

This chapter examines a different kind of meaning: one the philosopher Grice (1957) called "non-natural meaning" (henceforth: 'meaningNN'). Grice argued that meaningNN is expressed by behaviours that are, at least in the human domain, genuinely 'communicative', as opposed to ordinary instrumental behaviours like *pushing Roger out of the way*. It is important to keep in mind that, along with talking about a different notion of 'meaning', compared to earlier chapters, we are now also talking about a different kind of 'communication'. Recall Chapter 2, which weighed into the debate about what counts as animal 'communication'. There the relevant kind of communication was not restricted to individuals capable of intentional behaviour. In this chapter and the next, however, the focus is on a kind of communication (and the kind of meaning expressed and recognised in this kind of communication) that *is*

[57] This could be either in phylogeny or ontogeny.

restricted to intentional systems. However, it is restricted not only to animals capable of intentional behaviour, but also to animals that are, in addition, capable of *recognising* the intentions that lie behind others' behaviour. I will call this 'bi-intentional communication'. Henceforth, 'communication' will mean bi-intentional communication, unless otherwise indicated.

On the views I will be discussing in this chapter, communicative behaviour (i.e. meaningfulNN behaviour) is commonly distinguished from non-communicative behaviour (which doesn't express meaningNN) by the intentions giving rise to it. To meanNN something, one must behave motivated by an 'informative' intention, on the one hand, and a 'communicative' intention, on the other (Sperber & Wilson 1986). Moreover, communicative intent – the second kind of intention just referred to - is thought to include informative intent embedded within it. Because of this recursive intentional structure, signals carrying meaningNN are quite different from morphological features or behaviours carrying merely natural meaning. To carry natural meaning, as opposed to meaningNN, a behaviour need not be intentional, let alone recursively so.

The task of this chapter is to flesh out the notion of meaningNN, along with its explanatory commitments. The question will be: what psychological abilities do individuals need to possess to express and recognise meaningNN? The point of asking this question, for my thesis, is mostly comparative. MeaningNN is not, at least obviously, restricted to the communicative behaviour of human beings. While some researchers take the intentional demands of meaningNN to be a defining feature of human communication, and from this argue that human communication is set categorically apart from nonhuman communication (e.g. Scott-Phillips 2015), others argue that (some) nonhuman signals carry meaningNN: in particular, some great ape signals (e.g. Moore 2016; 2017). I am sceptical of these comparative claims, explaining why in the next chapter.

Section 1 begins with Grice's analysis of the concept of meaningNN, noting how this analysis was taken up by 'post-Griceans' interested in explaining the psychological abilities needed to express and recognise meaningNN. Section 2 explores the explanatory role of these intentions according to post-Griceans: more specifically, how the expression and recognition of these intentions constitutes the expression and recognition of meaningNN. Section 3 lays out the meta-representational abilities necessary for expressing and recognising meaningNN. This chapter is largely expository, laying the groundwork for the next one. As will become clear, the intentional demands of expressing and recognising meaningNN are significant. This presents a prima facie problem, seeing as young children can express and recognise speaker meaning[58] in the absence, arguably, of the recursive intentional abilities required for dealing in meaningNN (Moore 2017). As a result, Chapter 5 explores ways of deflating the intentional demands of human communication. It then examines whether nonhuman primates are capable of expressing and recognising meaning in same fundamental way humans do.

1. Grice's analysis

Grice sought a notion of 'meaning' that would apply to all and only those intentional behaviours that are 'communicative'. Intuitively, my pushing you out of the way is an intentional behaviour, but it isn't a communicative one. Intuitively, my asking you to go away *is* a communicative behaviour. Grice sought to reveal the necessary and sufficient conditions honouring intuitions like this about which intentional behaviours do and do not count as communicative. Genuinely communicative behaviours are those that carry meaningNN, according to Grice. Moore (2017) summarises what is required by Grice for a speaker S to meanNN something by uttering x to hearer H. I will take Moore's analysis as adequate in what follows. Specifically, S must utter x intending:

1. *H* to produce a particular response *r*
2. *H* to recognise that *S* intends (1).

The first condition for some behaviour to carry meaningNN is that it be intentional[59]. The content of the speaker's meaningNN is then identical to the effect the speaker intends their utterance to have on the hearer. Such ('informative') intentions can take a variety of forms. For example, a speaker might intend, by vocalising or gesturing, that the hearer do something: for example go away. Alternatively, the speaker might intend that the hearer believe something: for example the proposition *it's raining*. There are many other kinds of informative intentions a speaker might have in uttering. Each kind of informative intention corresponds to a particular type of effect a speaker intends his utterance to have on the hearer. Thus, for a speaker to meanNN something by behaving in a certain way, the speaker must intend something: typically, to modify the hearer's behaviour or mental states in some way. Conversely, for a hearer to recognise what a speaker meansNN, the hearer must identify the speaker's informative intention. This leads to a crucial point. Expressing and recognising

[58] Thus, it is an open question whether speaker meaning is in fact identical with meaningNN.
[59] The push-up display of a lizard is unlikely to be intentionally produced. Likewise, many think the alarm calls of vervet monkeys are not intentional (Seyfarth & Cheney 2003; Tomasello 2008). In the next chapter, exactly what intentional behaviour is will be explored in some detail as it is crucial to the issue of attributing meaningNN. This fundamental issue has, by and large, been ignored in the relevant literature.

meaningNN depends on the cognitive abilities that allow us to understand each other as intentional beings. To say that I meanNN something, it is necessary that I perform an intentional behaviour. Concomitantly, to say that Rick understands what I meanNN, it is necessary that Rick recognises the intention motivating my behaviour.

Behaviours that carry meaningNN were thought by Grice to be unique in a second important respect. They are not merely intentional, they are "overtly intentional" (Sperber & Wilson 1987). Imagine I am your boss. I could intend to bring it about that you leave your office by setting the fire alarm off. Alternatively, I could intend to bring it about that you leave your office by making you realise that I intend you to leave your office. The latter is meaningfulNN, while the first is not. Likewise, I could intend for you to buy me another drink by moving my empty glass into your field of view hoping that you notice my glass is empty and believe it's your turn to buy the next round. Alternatively, I could move my empty glass into your field of view while also making eye-contact with you, pointing at my empty glass, and raising my left eye-brow. In the second case (but not the first) it is plausible that, in addition to intending that you believe the next round is on you, I moreover intend that that you recognise I have this intention. In this way meaningNN is *overtly* intentional.

- Justification or description? –

Taken as an analysis of the necessary and sufficient conditions separating communicative behaviours from non-communicative ones, the above analysis might have some appeal. However, depending on one's view of psychological explanation, this appeal declines when the issue switches from analysis to explanation. The picture Grice held of psychological explanation becomes tantamount here.

Some (e.g. Grandy & Warner 2017) think that, for Grice, psychological explanation is normative as opposed to descriptive. That is, psychological explanation specifies how we *ought* to think, or think under 'optimal' conditions, as opposed to how we *in fact* think. If this view of Grice is correct, then an intentional description of some behaviour, for Grice, is ultimately an exercise in normative reconstruction. As a result, the recursive intentional structure deemed necessary and sufficient for meaningNN is not something that Grice thought speakers typically have *in mind* whenever they communicate. Instead, this intentional structure is something hearers attribute to speakers when the former need to *justify* the actions and beliefs caused by the speaker's utterance. If all this is true, then the recursive intentional structure emerging from Grice's analysis of non-natural meaning was not intended to be part of a cognitive explanation of how communication operates.

What about someone holding a descriptive, as opposed to normative, picture of cognitive explanation? That person might think Grice's intentional structure is of little scientific importance. On the other hand, a recently influential group of researchers have taken the intentional structure analysed by Grice to be central to (descriptively) explaining how human communication works. These 'post-Griceans' take Grice's analysis of the intentions necessary for communication as central to explaining how communication is both a) initiated, and b) used successfully, once initiated (Sperber & Wilson 1986; 2002, Tomasello 1998; 2008, Bloom 2000, Csibra 2010). These accounts also form the basis of much developmental and comparative research (Gomez 1994, Hurford 2007, Tomasello 2008, Scott-Phillips 2015, Moore 2016; 2017). I will have much more to say on the explanatory status of intentional states in Chapter 5, as the issue is crucial for assessing accounts of communication that rely so heavily on intentional states.

2. From analysis to explanation

Recall from Section 1 that meaningNN is thought to be limited to behaviours that are *overtly* intentional. Making one's informative intent overt is thought to be necessary to communication for at least two reasons. First, so that communication can be initiated. Unless the hearer recognises that I am attempting to communicate, as opposed to, say, moving my glass away from some ants crawling on the table, he might simply take my glass-moving behaviour to be of no concern. Second, unless the hearer recognises that I am attempting to communicate via my glass-moving behaviour, he will not invest the cognitive resources required to bridge the inferential gap between the information carried by my explicit action (i.e. my glass-moving behaviour), on the one hand, and my informative intention, on the other. After all, my glass-moving behaviour is rationally compatible with a wide range of possible informative intentions. If my interlocutor is to successfully infer a) the *kind* of informative intention underling my action (a request for more beer, as opposed to an invitation to emote at the style of the glass, for example), and also b) the *content* of this request (that my interlocutor buys me another drink, as opposed to depositing the empty glass back to the bar, for example), then he must make various 'pragmatic' assumptions that fill the inferential gap between glass-moving behaviour, on the one hand, and a request for another drink, on the other.

Now, human communication depends on a wide range of psychological abilities, only a subset of which include skills related to understanding intentionality (or the mental states that underpin intentional behaviours: *mindreading*). However, in recent years significant explanatory focus has been placed on the ability of human communicators to both express and recognise intentions. There are a couple of reasons for this. First, the role of mindreading in human communication is deemed important because of the commonly perceived gap between the mindreading abilities of human beings, on the one hand, and similar abilities in nonhuman primates. This gap, combined with the fact that nonhuman great-ape communication is greatly impoverished compared with human communication, has encouraged the view that the evolution of more powerful mindreading abilities was crucial to the evolution of human linguistic (and non-linguistic) communication (Hurford 2007, Tomasello 2008, Scott-Phillips 2015).

Second, human communication seems to operate in a way that is arguably entirely unique. Unlike communication in any other species, human communication seems to be possible in the absence of a 'code'. The concept of a communicative 'code' is a not entirely a transparent one, but one of the main features of this concept is *meaning in virtue of 'convention'* (Lewis 1969, Skyrms 2010). Whether the convention is established by blind trial and error (either natural selection of simple learning), or via more sophisticated means (rational agency), and moreover whether the code involves a combinatorial structure or not, the resulting code fixes meaning in advance. Interlocutors must choose from a set menu of communicative options. In contrast, human communication is flexible to a degree that indicates that it goes above and beyond a code, even considering the combinatorial structure (i.e. syntax) of the codes *we* use. We employ codes (paradigmatically, languages) in communicating with each other, but we can make these codes *mean* a wide range of things. Meaning is not fixed in advance. Instead of codes determining meaning, they merely provide evidence of what we mean. Moreover, we can wield our linguistic codes flexibly, and at will. While it may take many generations of sex and death for a signal in a relatively simple communication system (say, frog mating calls) to acquire a different coded content, I can employ the coded content "do you want to get out of here?" to mean completely different things, depending on my intentions and the context. This is just to express the rather uncontroversial point that, when it comes to human communication, speaker meaning is not exhausted by coded (or 'semantic') content. However, this point has very significant implications for proponents of the post-Gricean picture. This is (and here we return to the issue of mindreading) because bridging the gap between coded meaning and speaker meaning is thought to rely on the ability of human beings to understand each other as intentional agents.

2.1 The Explanatory Role of Informative Intentions

The post-Gricean picture begins with the following proposition: in communication, the coded content of an utterance usually, if not always, underdetermines what the speaker *means* by uttering it. Coded content merely provides evidence for what the speaker means, as opposed to exhausting meaning. For example, the coded content of my utterance 'It's your turn', said to my friend Rick at drinks, represents[60] that it is his turn to do something. But exactly what it is his turn to do is not specified. Thus, for communication to be successful, Rick must somehow bridge the gap between the coded content of my utterance, on the one hand, and the *intention* (or goal) I have in producing this utterance, on the other. In the present example, my intention is that Rick buys the next round of drinks. To infer my intention (which is a necessary step to complying with it), Rick must move from representing the coded content 'It's your turn', to representing my goal that he believes it is his turn to buy the next round. By doing so Rick grasps what I *mean* (as opposed merely to the coded content of the sentence I uttered).

Thus, as mentioned, there is a clear distinction being made between coded content, on the one hand, and speaker *meaning*, on the other. The former is simply the conventional content of the linguistic code, and is a product of combining lexical items according to grammatical rules to produce a semantic representation: but one which, by itself, often (if not always) underdetermines speaker meaning. The second thing to note is the way in which the gap between coded linguistic content, on the one hand, and speaker meaning, on the other, is inferentially bridged (by the hearer). Sometimes, bridging this gap will no doubt involve the application of relatively simple heuristic rules. At other times, however, it involves the hearer running a rich and flexible online model of what is "mutually manifest" (Sperber & Wilson 1986) or "common ground" (Tomasello 2008) to both speaker and hearer.

An example will help illustrate the phenomena of mutual manifestness/common ground[61]. Imagine that Rick has finally bought the next round of drinks and we are once again both sitting at our table chatting away. We happen to be sitting next to each other, and I am on the right-hand side of Rick. Suddenly, mid-conversation, I happen to randomly glance over my right shoulder. As a result of doing this, I spot Roger, an associate who we both do not particular like, walking towards us. I lean back suddenly so that Rick can see that Roger is coming, with eye brows clearly raised at Rick. My actions (in particular, my leaning back) meanNN something like 'Arrgg, here comes Roger, prepare yourself!'. My intention in leaning back is to get Rick to believe that Roger is coming, and to ready himself for an annoying exchange with a person he doesn't like. However, in order that Rick infers my intention (my *meaning*), he must understand what I will *expect* him to attend to in his newly modified visual field. Only by representing the kinds of things that are thought by me to be salient to him will Rick be able to 'connect the dots' between my

[60] I use the term 'represents' as opposed to 'means' because I am restricting the notion of '(non-natural) meaning' to speaker meaning, as opposed to semantic/coded content.
[61] I will speak of 'mutual manifestness' in what follows, but I do not thereby intend to be choosing one account over the other: both the idea of mutual manifestness and of common ground are close enough to each other for present purposes.

leaning back, on the one hand, and my intention (my meaning) in so acting, on the other. For there are multiple sources of new information that Rick might attend to because of my leaning back: the bar window, the weather outside, various people and tables, someone he finds attractive, and anything else that my body was formally obstructing from his view.

This is where the notion of 'relevance' enters, at least on one post-Gricean account (Sperber & Wilson 1986). Because on this account cognition is 'relevance-based', then if I have an internal model of what is mutually manifest to both Rick and myself, I can have a decent (if fallible) handle on what assumptions Rick is likely to make in trying to get from my utterance, or communicative act (i.e. my leaning back), to my informative intention or meaning. This raises the question of what it means to say that cognition is 'relevance-based'. The hypothesis is that cognition seeks to bring about the greatest cognitive upshot for the least amount of processing effort. What is cognitive upshot? Sperber & Wilson (1986, p. 48) posit that it is the derivation of inferences from new information. On this view, what makes a potentially new bit of information *p* more *relevant* than *q*, is that *p* can be combined with old information (i.e. in memory) to give rise to the derivation of more inferences than would follow from *q* being combined with this old information (assuming processing effort is equal)[62].

A plausible assumption of the view is that conceptual processes can't attend to all stimuli made available by the sense organs at any moment. This is reflected in the fact that, when I lean back for Rick, he cannot attend to everything in is newly modified visual scene. But because human cognition operates according to the principle of relevance, Rick will attend to that feature of the visual scene which, given the information already contained in his mind, will give rise to the most inferences. When it comes to our imaginary scenario, what gives rise to the most inferences in Rick's mind is the sensory information indicating the approach of the malign Roger. Thus, this information is most *relevant* to Rick. From this new information (i.e. that Roger is approaching), Rick will be able to derive a host of new information, such as that he should move, or avoid behind seen, or prepare for a sarcastic exchange. Given the information already stored in Rick's mind, including the belief that he finds Roger intolerable, much more follows (for Rick) from new information about the approach of Roger, than, for example, new information about the current state of the bar window (also revealed by my leaning backwards).

What does this have to do with communication, and with mutual manifestness, in particular? The idea is that if I know what information is potentially

---

[62] One might wonder if a circularity lurks here. Cognitive upshot, in any useful sense of that term, surely can't be measured by the quantity of new information inferred from old information. This is because an enormous number of uninteresting facts can be inferred from just about any additional item of information. Instead, cognitive upshot must surely consist in the quantity of *salient* (i.e. relevant!) new information inferred from old information.

manifest to Rick in the environment I share with him, and if I 'know' (if only implicitly) that his mind seeks *relevant* information, then as a result I have a fairly decent idea of what stimuli to produce for him in order that he is likely to infer what I want him to. To bring it back to the imaginary example, I know that Rick dislikes Roger and that Rick will find the approach of Roger salient. Thus, by leaning back on my chair and creating a clear line of sight between Rick and the impending Roger (as well as other, less relevant things), I can be quite confident that he will infer my intention in so acting (i.e. that he believe roger is approaching, as opposed to, say, believing the weather outside is nice).

A slightly different variant on the standard picture[63] focuses on something known as the 'cooperative principle' as opposed to the principle of relevance, but the broad moral is similar. The cooperative principle is a normative claim - about what interlocutors *should* do when communicating - as opposed to a descriptive claim - about *how* the psychological processes underpinning communication *do* operate. However, as with the (descriptive) principle of relevance, the (normative) cooperative principle provides a way to 'plug the gap' that exists between literal-linguistic ('coded') meaning, on the one hand, and speaker meaning, on the other. Just like the principle of relevance, the cooperative principle holds that "communication creates expectations which it then exploits" (Sperber & Wilson 1986, p. 37).

The cooperative principle states that speakers and hearers should try to adhere to certain guidelines when performing their respective roles in an episode of communication. From knowledge that the speaker is following certain communicative norms, it should be possible for the hearer to infer the speaker's meaning from observation of a) the speaker's utterance/action and b) the context. This is because, out of all the possible thoughts a speaker might have in mind by uttering, for instance, 'It's your turn!', only one will be compatible with the cooperative principle: namely, 'It's you turn *to buy the next round of drinks*'. Thus, the speaker's role is to aim to produce an utterance which will have only one interpretation compatible with the cooperative principle (just like, above, the speaker is simply calibrated – as a matter of psychological design - to produce an utterance with only one interpretation compatible with the principle of relevance).

The cooperative principle states a set of 'maxims': for instance, of quantity, quality, relation and manner. These include norms like a) make your contribution as informative as required without giving too much information, b) do not say what you believe to be false or what you lack sufficient evidence for, c) avoid ambiguity, e) be relevant, and so on. If the hearer supposes that the speaker is following such maxims according to the cooperative principle, then he can eliminate those interpretations of

[63] Indeed, one that is more directly indebted to Grice than is relevance theory, the latter being usually referred to as a 'post-Gricean' theory of communication (Scott-Phillips 2015).

the speaker's utterance incompatible with the idea that the speaker is following the maxims. For instance, me overtly leaning back on my chair is relevant to Rick only if he assumes that I intended him to observe the impending Roger.

In summary, informative intentions are the 'stuff' of communication. They constitute what I (non-naturally) *mean* in communicating with you. The hearer's 'job', in communicating, is to identify the speaker's informative intention. One important issue to be explained, in understanding communication, is how the hearer manages to grasp the speaker's informative intention, given that coded meaning usually (if not always) underdetermines the speaker's informative intention. Furthermore, successful communication often occurs in the complete absence of a coded utterance (as when I lean back for Rick, or when I'm travelling in a foreign country and don't speak the native tongue). The standard picture posits that the hearer inferentially grasps speaker meaning by interpreting the actions of the speaker in accordance with a key assumption: that the speaker's utterance was produced in line with the principle of relevance (or something functionally similar). As a result of operating in accordance with this assumption (and if the speaker has indeed fulfilled his part of the bargain), it should be possible for the hearer to infer something close enough to the speaker's informative intention by taking into account any coded content or other action produced, along with information about what is mutually manifest to both speaker and hearer.

2.2 THE EXPLANATORY ROLE OF COMMUNICATIVE INTENTIONS

This leads us to the topic of communicative intentions. According to one interpretation of the Gricean analysis of meaningNN (i.e. Moore 2016), an action is meaningful or communicative (as opposed to just another ordinary, non-communication, action) only when the speaker intends that his informative intention is manifest to the hearer. That is, communication involves (1) the speaker intending his utterance to have effect *r* on the hearer (informative intention), *along with* (2) intending that (1) is manifest to the hearer (communicative intention).

Why is this important to the process of communication? Firstly, so that a hearer knows when an action is intended to be communicative, as opposed to non-communicative. Unless Rick thinks that by leaning back on my chair I intend to change his beliefs about the world, he won't have any reason to make the relevant inferences about my informative intention in so leaning back. In other words, communicative intentions *signal signalhood* (Scott-Phillips 2015). This has a secondary importance, related to the issue of relevance. In Section 2.1, we explored the role that relevance plays in communication according to influential post-Gricean accounts. Communication involves a hearer assuming (if implicitly) that the speaker's utterance was produced according the principle of relevance (or something similar). Now, the role of communicative intent is precisely to signal when actions are, in fact, produced in this way. That is, communicative intentions are intentions that the hearer recognise that the speaker has produced a stimulus, for the hearer, that is guaranteed to be relevant.

This can be seen mostly vividly when communication involves no coded content, such as when I lean back for Rick. Recall that in the hypothetical scenario my action (i.e. leaning back) was accompanied by another: me raising my eyebrows with face clearly oriented towards Rick. The eyebrow raise is me signalling signalhood to Rick. It is the result of my communicative intention that Rick recognise that my leaning back is relevant for him. My expressing such an intention and Rick recognising such an intention is crucial so that he invests some cognitive resources in identifying my informative intention in leaning back. Because my eyebrow-raise signals signalhood - and with signalhood, relevance - Rick can discount non-relevant stimuli that I have drawn his attention to by leaning back. After doing so, he will be left with the information that Roger is looming, and that it was my (informative) intention to make him believe this. In short, Rick is in a much better position to infer my informative intention from having recognised my communicative intention.

To make the role of my communicative intention even more vivid, imagine that roger is approaching the bar from outside and is some distance away, mixed in with other people on the move, and other features of the environment, such as the car park, people sitting on outside tables, and so on. In this case it will be even more essential that I signal communicative intent to Rick, in order that he might accurately infer my informative intention. In the former case, where Roger was approaching our table from inside the bar, my act of leaning back might have alerted Rick to Roger. However, in this case were Roger is approaching from outside the bar at some distance and mixed in with other people, only by indicating to Rick that there is something of relevance for him to attend to in that direction will he invest the cognitive resources needed to distinguish the important information from all the other distracting information.

It is crucial to note that, sometimes, communicative intentions will be expressed in a separate action to the one expressing the informative intention, as in the imaginary example we have been entertaining so far (the eyebrow raise). However, they need not be. Often, communicative intentions can be expressed using same action as the utterance, such as when I cry 'Look out!', after badly slicing a golf ball in the direction of some unfortunate bystanders. In this case, my guarantee of relevance is attached to the cry itself. Indeed, all cases of linguistic communication are thought to be accompanied by a tacit guarantee of relevance, explaining why we find the person who goes on and on in a conversation annoying. He is providing far too much information, violating the principle of relevance. We are automatically drawn to expect relevant information - new information which can combine with information we already possess to derive new information - when, to our displeasure, there simply is none.

In summary, the role of the communicative intention is to signal signalhood, and thus relevance (or alternatively, the fact that the speaker is operating according to the cooperative maxims, or something functionally similar). I should stress that I do not necessarily endorse this account of communication. I am simply trying to make the account explicit to reveal its cognitive commitments. In the next section, I address what some take to be the cognitive abilities needed to express and recognise both informative and communicative intentions.

3. The psychology of expressing and recognising meaningNN

According to recent post-Gricean explanatory accounts of communication[64], the cognitive abilities required to express and recognise meaningNN are substantial. I lay them out in this section. I split the performance analysis into two parts: initiating communication (expressing and recognising communicative intent), on the one hand, and using communication successfully (correctly producing and inferring informative intentions), on the other.

3.1 INITIATING COMMUNICATION

- Speakers -

1st-order intentionality: A speaker must possess at least a 1st-order intention towards a hearer. In the case of directive communication, my informative intention is an instance of 1st-order intentionality: i.e., I <u>intend$_1$</u> that you *go away*.

2nd-order intentionality: Sometimes, however, we communicate with the intention that H *believe* (or take some other cognitive attitude towards) some proposition *p*. That is, sometimes we are not communicating with an intention to merely influence behaviour (as with gesturing you to go away, for instance). Rather, I might (as my informative intention) <u>intend$_1$</u> that you <u>believe$_2$</u> (for example) that *it's raining*.

[64] Sperber 2000, Tomasello 2008, Scott-Phillips 2015. Sperber and Scott-Phillips, as opposed to Tomasello, are quite explicit about the psychological abilities required. However, Tomasello does sometimes explicitly state that multiple orders of recursive mindreading are necessary for human communication (e.g. Tomasello 2008, p. 159).

3rd and 4th-order intentionality: We are yet to introduce the communicative intention into the picture. In order to signal signalhood, and so prepare you for relevant information, I must (according to the standard post-Gricean picture) <u>intend</u>$_1$ you to <u>believe</u>$_2$ that I <u>intend</u>$_3$ you to <u>believe</u>$_4$ *it's raining*. In order that you realise I have intentions towards you produced in accordance with the principle of relevance (or something similar), I must not only intend you to be influenced in line with my informative intention. In addition, I must also intend you to recognise I intend to affect you in this way. Only then will you invest the required cognitive resources needed to infer my informative intention, given the under-determination of my informative intention by coded linguistic meaning (or other signals). Alternatively, if the speaker's informative intention is merely that the hearer go away, as opposed to believe (or doubt, etc.) a proposition, then S need only <u>intend</u>$_1$ that H <u>believe</u>$_2$ S <u>intends</u>$_3$ H to *go away*.

- Hearers -

2nd-order intentionality: Take a speaker S motivated by the informative intention that H go away. In order for H to grasp S's meaning (i.e. S's informative intention), H must <u>believe</u>$_1$ S <u>wants</u>$_2$ H to *go away*.

4th and 5th-order intentionality: Now factor in S's communicative intention. That is, assume S utters intending his informative intention to be manifest to H. In order for H to grasp this, H must <u>recognise</u>$_1$ that S <u>intends</u>$_2$ H to <u>believe</u>$_3$ that S <u>wants</u>$_4$ H to *go away*. More demanding still, assume now that the speaker is motivated by the informative intention that H believe something. For H to grasp S's communicative intention in this situation (i.e. where S has an informative intention directed at H's mental states and not just her behaviour), H must <u>recognise</u>$_1$ that S <u>intends</u>$_2$ H to <u>believe</u>$_3$ that S <u>wants</u>$_4$ H to <u>beleive</u>$_5$ (say) *it's raining*.

In all, according to this performance analysis, initiating communication requires at least 3rd-order intentionality on the part of speakers, and at least 4th-order intentionality on the part of hearers (Sperber 2000, Scott-Phillips 2015). But this is when communication is directed at behaviour. When the goal of communication is to get the hearer to believe something, for example, then communication requires at least 4th-order intentionality on the part of speakers and 5th-order intentional on the part of hearers. We now turn to the issue of using communication successfully, once initiated.

3.2 Using communication successfully

According to Sperber & Wilson (1986) and Tomasello (2008), recursive mindreading is needed to ensure that communication proceeds successfully, once initiated. More specifically, it is needed so that speakers provide the right evidence for their informative intention, and so that receivers make the appropriate inferences to arrive at the speaker's informative intention (or something close enough). This is essential to the notion of 'mutual manifestness' (Sperber & Wilson 1986) or 'common ground' (Tomasello 2008). The idea is that a speaker must represent what is mutually manifest or common ground between the hearer and himself.

Turning back to our example, I must provide the right evidence for my informative intention, given what I believe about what information is manifest to Rick, along with what subset of this information will be relevant to Rick. Unless I know that, upon leaning back, Rick will be able to see Roger, and moreover that this information will be salient for Rick, I won't know whether leaning back will be sufficient to fulfil my informative intention. That is, I must <u>think</u>$_1$ that Rick <u>see and attend</u>$_2$ to Roger, upon leaning back on my chair. Concomitantly, Rick must be able to infer my informative intention, given what he believes - about what I believe - he will find salient. Unless Rick knows that I think he is able and willing to attend to Roger as a result of my leaning back on my chair, he won't be confident in inferring my informative intention. That is, Rick must <u>believe</u>$_1$ that I <u>think</u>$_2$ he will <u>see and attend</u>$_3$ to Roger, upon me leaning back on my chair.

Arguably, it doesn't stop there. It seems, further, that I must have some understanding that Rick has this nested awareness of what I think he will see and attend to, in virtue of me leaning back on my chair. Otherwise, why would I lean back on my chair, as opposed to provide some other kind of evidence for my informative intention? That is, unless I <u>think</u>$_1$ Rick will <u>believe</u>$_2$ I <u>think</u>$_3$ he will <u>see and attend</u>$_4$ to Roger, upon me leaning back on my chair, it won't be rational to lean back on my char. If we stop here[65], it appears that representing what is mutually manifest or common ground is not trivial, by any means. It requires at least 3 orders of intentionality for hearers and 4 orders of intentionality for speakers. In the next chapter I largely ignore the issue of using communication successfully, once initiated. I focus exclusively on the intentional demands of initiating communication. This issue has received the most attention in recent literature: probably because of the intimate connection between Grice's intentional requirements for an action to count as meaningfulNN, presented in Section 1, and the role of the communicative intention (Grice's first intentional requirement) in initiating communication, as opposed to allowing it to proceed successfully once initiated. This isn't to say the issue of successfully using communication is not important. On the contrary, if the performance analysis presented here is on the right track, then proponents of the Post-Gricean account will need to show how the relevant intentional requirements of

[65] No doubt there are some that would be inclined to expand this recursive dance of intentional states out even further.

successfully using communication are met by young children in the face of the development findings on mindreading abilities presented in Chapter 5[66].

4. Conclusion

If the post-Gricean picture is correct, then the kind of meaning humans express and recognise is psychologically demanding. In the next chapter, we will see that it is arguably *too* demanding, given the relatively modest mindreading abilities of young children who are nevertheless learning to express and recognise speaker meaning. The focus then goes back to animal communication. If the kind of meaning humans trade in must be deflated, then this has clear comparative implications. The simpler human communication is, the more likely other species are doing it, or something close to it.

*   *   *

Chapter 5. *What of young children able to express and recognise speaker meaning in the absence of explicit higher-order mindreading abilities? Perhaps this mindreading is done implicitly, below the level of conscious awareness and explicit control? Or perhaps humans typically express and recognise a fundamentally simpler kind of meaning than Gricean non-natural meaning? What are the comparative ramifications? What kind of meaning do chimpanzees and monkeys send and receive?*

[66] It should be noted that I have presenting a simplified version of the Sperber & Wilson line. For instance, they have offered specific proposals for how speaker meaning is coded and decoded.

Chapter 5

Grades of Speaker Meaning

This chapter addresses an important objection to the post-Gricean picture (Moore 2017). Young children are able to express and recognise speaking meaning in the absence of explicit higher-order mindreading abilities. In response to this objection, aired in Section 1, I present some options in Section 2. One option I present is to drop non-natural meaning, and with it, Grice's analysis as being an accurate depiction of the kind of meaning humans *typically* traffic in. As a result, I propose kinds of meaning standing in-between natural meaning, on the one hand, and meaningNN, on the other. Inspired by Moore (2016; 2017), but also parting ways with him on some of the important details, I draw on the work of Csibra (2010) to first propose a kind of speaker meaning ('meaningC'), in which speakers behave intending their communicative behaviours to effect hearers in a particular way; and where hearers must grasp something close enough to the speaker's intention to accurately infer speaker meaning. What is not required is the recognition of communicative intent. Instead, the functional role of recognising communicative intent is handled by signals carrying *natural meaning* about the sender attempting to communicate, which are sent and received without meta-representation. I also propose a third notion of speaker meaning (in addition to meaningNN and meaningC), involving no proxy for the expression and recognition of communicative intent whatsoever. Finally, Section 3 returns to animal communication. There, I argue that chimpanzees don't meanNN things when gesturing. However, they may well express and recognise a more modest form of speaker meaning. I then argue that the pragmatic-inferential abilities of old-world monkeys do not show them to expressing and recognising any kind of speaker meaning, despite these inferential abilities representing an important step on the road towards the domain-general nature of human comprehension. In short: old-world monkeys are 'simply' more sophisticated interpreters of natural meaning, while chimps may send and receive a form of speaker meaning: i.e. a kind of meaning requiring intention expression and recognition.

1. Ontogenetic constrains on post-Gricean communication

In Chapter 4, I presented a broad sketch of a currently-influential explanatory account of human communication. In this account, the dual intentional structure analysed by Grice is put to explanatory work[67]. Speakers have higher-order intentions towards hearers, and hearers must represent these higher-order intentions. This is both so that communication can be initiated and so that it can unfold successfully, once initiated. In this section, I explore ways to deflate the intentional demands of the standard post-Gricean picture, in response to an important objection (Moore 2017).

To begin with, recall again the conditions for some act to have meaningNN (Moore 2016). A speaker (or non-verbal communicator) S non-naturally means something by an utterance x if and only if, for some hearer (or audience) H, S utters x intending:

1. H to produce a particular response *r*, and
2. H to recognise that S intends (1)

Unless Rick recognises that I *intend* him to believe something, he might take my leaning back as just an ordinary piece of instrumental behaviour on my part. Perhaps I'm just adjusting my posture to relieve some discomfort. Or perhaps I'm shifting my posture to get a better look at someone across the room. The role of clause 2 is to signal signalhood, and with signalhood, relevance. By leaning back I might intend to produce a particular response in Rick, namely the belief that Roger is approaching (clause 1). But unless Rick interprets my leaning back as communicative, Rick will be less likely, according to the standard picture, to pragmatically infer my informative intention from my rather indeterminate outward behaviour. To this end, I must intend Rick to recognise that I intend him to believe Roger is approaching.

One objection grants the necessary role of clause 1 but denies that of clause 2, at least in the form presented above. In leaning back, I might plausibly [$\underline{\text{intend}}_1$ Rick to $\underline{\text{believe}}_2$ Roger is approaching]. But why must I $\underline{\text{intend}}_1$ that Rick $\underline{\text{recognise}}_2$ [I $\underline{\text{intend}}_3$ him to $\underline{\text{believe}}_4$ Roger is approaching]? The objection is that, in order to signal signalhood, I need not possess a communicative intention within which [the content of my informative intention] is embedded. Instead, my communicative intention can take a simpler form. Moore (2016; 2017) argues for this possibility. He does so for good reason. For if the performance analysis presented in Chapter 4 was correct, then arguably young children should be unable to begin communicating. But young children *do* begin communicating. In Chapter 4 we saw that, on the standard picture, initiating communication requires at least 3 orders of intentionality of the part of

[67] As mentioned, Chapter 6 examines the status of intentional psychology in detail.

senders and at least 4 on the part of receivers. The presents a potential problem. The problem is that children seem to be pragmatically competent communicators well before they develop the ability to process 3 or 4 orders of intentionality. Two findings are in prima facie conflict: First, experiments carried out on children suggest that only around ages 5 to 6 do they pass (explicit) second-order false belief tests[68], let alone tasks requiring more sophisticated meta-representation (Miller 2009). At the same time, however, the communicative behaviours of children as young as 12 months suggest that they can take into account their own past experiences with an interlocutor (Liebal et al. 2009) and are motivated by what appears to be something as rich as communicative intent (Grosse et al. 2010[69]).

At least one proposed solution to this prima facie conundrum is to assert that young children are engaging in *implicit* higher-order mindreading. I will address this proposal soon in Section 2.1. Another proposal, mentioned already, is to reduce the demands of post-Gricean communication. Moore (2017) takes this second route, arguing that communicative intentions need not be as complex as post-Griceans think. For instance, Scott-Phillips (2015, p. 67-68) adopts the standard account of Sperber & Wilson, claiming that "the content of a communicative intention is an informative intention". Communicative intentions are <u>intentions</u>$_1$ that hearers <u>recognise</u>$_2$ that [speakers <u>intend</u>$_3$ hearers to do/<u>believe</u>$_4$ X], where the informative intention in brackets is part of the content of the communicative intention. In contrast to this, Moore argues that communicative intent can be functionally distinct from informative intent. This simplifies communicative intent by removing some of its recursiveness. We have a distinct communicative intention, on the one hand, and a distinct informative intention, on the other. What might be the content of functionally-distinct, 'Moorean', communicative intent? Moore (2017) suggests the content could be as simple as *S <u>intends</u>$_1$ that H respond to S's signal*. He argues that such an intention might give rise to 'acts of address' performed before signalling. Acts of address are behaviours like moving into H's line of sight, making eye content with H, or calling H's name. For instance, before leaning back on my chair I make eye contact with Rick and raise my eyebrow. This is a perfect example of an act of address. On Moore's proposal, then, the following functionally-distinct intentions suffice for successfully initiating communication:

1. I <u>intend</u>$_1$ (in making eye contact and raising my eyebrow) that Rick respond to my impending behaviour (i.e. my leaning back on my chair)
2. I <u>intend</u>$_1$ (in leaning back on my chair) that Rick <u>believe</u>$_2$ Roger is approaching

[68] Passing a second-order false-belief tests shows that the subject can process 3rd-order intentionality: e.g. I <u>represent</u> X's <u>representation</u> about Y's (or my own) <u>representation</u>.
[69] Outlined in more detail in Section 2.1.

Intention 1 is my (Moorean) communicative intention. Intention 2 is my informative intention, or my intended message. We can further deflate the intentional demands of the example by changing things slightly, such that my informative intention is imperative in nature, as opposed to informative:

> 2a. I intend$_1$ (in leaning back on my chair) that Rick turn to the direction previously obstructed by my body

The result is that speakers need only 1st-order intentionality to initiate simple forms of communication. Hearers, for their part, will need only 2nd-order intentionality to represent the two 1st-order intentions of the speaker. This is good, as empirical evidence suggests children as young as 15 months have an implicit (i.e. non-linguistic) understanding of others as representing the world (Onishi & Baillargeon 2005) or as being motivated by a goal (Gergely & Csibra 2003)[70].

There is reason to think Moorean communicative intent is too simple, however. If communicative intent is as simple as Moore suggests, then successful communication would be initiated far less frequently than it in fact is. Recall the importance of signalling signalhood, on the post-Gricean picture. Moorean communicative intent will need to motivate behaviour(s) that robustly get Rick to realise that my leaning back on my chair has the function of communicating something, as opposed to (say) relieving back strain. But the intention that Rick merely *respond* to my main signal will not suffice. If, in making eye contact and raising my eyebrow, I am motivated by something as simple as the intention that Rick respond (but in what way?) to my leaning back, then I will be satisfied if Rick thinks, by leaning back, I am merely stretching. Of course, my informative intention wouldn't be satisfied in this scenario, but my (so-called!) communicative intention would be, despite the fact I've failed to initiate communication. Acts of address have a function within the post-Gricean framework, and this function is to signal signalhood to the hearer. If acts of address are to typically perform their function robustly, they will need to be produced with a more specific intention in mind than that H simply respond to S's signal. They would have to be produced, typically, by S intending that H respond to S's signal *in a particular way*. S must intend$_1$ something like the following: that H recognise$_2$ S is communicating[71] (as opposed to, say, think$_2$ that S is stretching – or doing something

[70] Heyes (2014, 2015) asks whether what is called implicit 'mindreading' in such cases is not actually mindreading, i.e. the representation of others' internal mental representations, but rather what she calls 'submentalizing': (implicit) cognitive processing enabling social coordination without the agent representation others' mental states.

[71] Here 'communicating' can be replaced by 'is doing something of relevance to me', or something functionally similar.

else). H would then need to recognise$_1$ that S intends$_2$ H to recognise$_3$ S is communicating.

If this criticism is fair, then we are left in the following position. The original motivation for splitting the communicative intention from the informative intention was to accommodate the lack of empirical evidence that children can represent something as recursive as Gricean communicative intent: i.e. my intention$_1$ that Rick recognise$_2$ I intend$_3$ him to look in such-and-such a direction. Instead, Moore suggests the expression and recognition of the following functionally-distinct intentions suffice:

1. I intend$_1$ (in making eye contact and raising my eyebrow) that Rick respond to my impending behaviour (i.e. my leaning back on my chair)
2. I intend$_1$ (in leaning back on my chair) that Rick turn to the direction previously obstructed by my body

Against this, I have just argued that the following functionally-distinct intentions are in fact required:

3. I intend$_1$ (in making eye contact and raising my eyebrow) that Rick recognise$_2$ I am communicating with him
4. I intend$_1$ (in leaning back on my chair) that Rick turn to the direction previously obstructed by my body

If communicative intention 3 is indeed required, as opposed to Moorean communicative intention 1, then speakers will need two orders of intentionality to communicate and hearers will need at least three. Hearers will need to recognise$_1$ that S intends$_2$ H to recognise$_3$ that S is communicating. If this is true then the conundrum introduced in this section stands: young children seem capable of a form of communication requiring at least 3rd-order intentionality (to recognise communicative intent), yet empirical psychology shows that, at present, children must reach 5 or 6 years of age before then can (at least explicitly) process three orders of intentionality[72] (Miller 2009). Thus, the cognitive demands have been cut down: there is no need for 4th or 5th order intentions. However, the demands are still too much for toddlers, prima facie.

2. Some options

[72] Or what Miller calls "second-order mental-state attribution": as already mentioned, the subject's ability to represent X's belief about Y's (or the subject's own) belief.

As I see it there are at least three possible solutions. The first is the most radical in the present context. It involves discarding the post-Gricean account, *qua* psychological explanation of communication. If hearers need at least 3 orders of intentionality for communication to be robustly initiated, and if children can understand utterances before they have developed the ability to recognise 3rd-order intentionality, then something must give. Perhaps the Gricean framework is best viewed as philosophy rather than science: i.e. as an account of how speakers and hearers would ideally *justify* their communicative behaviours, if pressed (Grandy & Warner 2017).

A more conservative second option, at least from the perspective of the standard picture, is to explore the possibility that the meta-representational demands of expressing and recognising communicative intent are managed implicitly, by System 1 cognitive processes. This route is taken by Scott-Phillips (2015). He highlights that tests of higher-order mindreading in children have been restricted to tests of *explicit* higher-order mindreading. As a result it might be the case that, as with 1st-order mindreading skills, children possess the cognitive abilities to pass *implicit* testing before they pass explicit testing (Onishi & Baillargeon 2005).

There is a third option. It mirrors the second in hypothesising that the coordination required to initiate communication is enabled by System 1 cognitive processes. However, unlike the second option, here System 1 processes merely mimic the expression and recognition of recursive communicative intentions. They do not process implicitly represented recursive intentional states. This is how I interpret Csibra (2010). In what follows I explain options 2 and 3 in more detail. Each gives rise to different empirical predictions, which is hopefully a virtue of distinguishing them. For what it's worth I argue that option 2 deserves to be considered 'Gricean'/'post-Gricean' in spirit, while option 3 does not.

2.1 SYSTEM 1 OPTION A: IMPLICIT HIGHER-ORDER MINDREADING

A strategy for those wishing to accommodate human communication within a broadly Gricean[73] framework is to push the meta-representational demands attached to expressing and recognising communicative intent below the level of conscious awareness and executive control. The idea is that speakers need not explicitly infer what behaviours are likely to initiate communication. Likewise, and in response to

[73] Unless otherwise indicated, by 'Gricean' I will from this point forward mean a form of communication, indebted to Grice, but where (unlike with both fully-Gricean and post-Gricean communication) communicative and informative intentions are functionally-distinct mental states, following Moore. As I have argued though in Section 1, and in parting ways with Moore, the former kind of communication still requires recursive mindreading on the part of speakers and hearers: hence the two System 1 proposals currently being considered. I will use the term 'post-Gricean' to refer to the standard account examined in Chapter 4, involving upwards of 3rd-order intentionality for speakers and 4th-order intentionality for hearers, which was then deflated (but only somewhat) in Section 1.

such behaviours, hearers need not explicitly infer that speakers want to communicate. Instead, the relevant inferential work is done implicitly.

The difference between explicit and implicit varieties of inference is presented by dual-processing theories of cognition (Evans 2008, Frankish 2010). Dual-processing theories take there to be two fundamentally different processing 'modes' available for many cognitive tasks: one that is fast, automatic and non-conscious; another that is slow, controlled, and conscious. The first (implicit) mode is relatively effortless and runs in parallel to tasks placing demands on reflective cognition. In contrast, the second (explicit) mode requires significant cognitive effort and can be disrupted by (other) tasks placing demands on explicit cognition. The first mode is labelled 'type 1' processing; the second 'type 2'. Dual-systems theory extends these findings, assigning the two different processing modes to two separate cognitive systems: 'System 1' and 'System 2', respectively. System 2 is often taken to be uniquely human, and the source of our ability to think abstractly and hypothetically.

The hypothesis for this section is that communicative intent is processed implicitly, by System 1. This hypothesis is at least consistent with introspection. It certainly *seems* like acts of address, underpinned by communicative intent, are not performed and recognised explicitly. If you pass me in the hall and make eye contact with me before nodding hello, to direct my communicative attention into your nod, you are probably not consciously reasoning thusly: "I desire that David interpret my nod as communicative. I believe that, for David to interpret my nod as communicative, I need to direct his attention onto it by making eye contact with him before nodding. Therefore, I will make eye contact with him before nodding". Likewise, it doesn't seem like I need to explicitly represent your recursive intention that I interpret your eye contact as evidence of your intention that I recognise you are trying to communicate with me. Instead, the expression and recognition of communicative intent seems to be carried out implicitly.

Two versions of this proposal can be distinguished, the first to be addressed here and the second in Section 2.2. On the first version of this proposal, I possess a mental representation of your communicative intention, the content of which is recursive. However, my mental representation of your communicative intention is not subject to conscious awareness nor executive control: it is processed in System 1. Nevertheless, the content of this representation is isomorphic to an explicit representation taking the following form: *S intends*$_1$ *me to recognise*$_2$ *that S is communicating* (or something close enough). This version of the proposal is suggested by Scott-Phillips (2015), who is sensitive to the prima facie conflict between current empirical evidence on explicit false-belief tests, on the one hand, and the intentional demands of post-Gricean communication, on the other. Children pass explicit false-belief tests only around 4 years of age, and this finding is robust (Wellman et al. 2001)[74]. Scott-Phillips argues that the post-Gricean account is saved in virtue of implicit, 1^{st}-order, false-belief testing (e.g. Onishi & Baillargeon 2005) showing that children possess non-verbal mindreading skills as young as 12 months. However, the proposal places a significant empirical bet that might not be borne out. It requires that implicit (i.e. System 1) mindreading tests one day scale up to show that we possess implicit *recursive* mindreading abilities: i.e. that System 1 mindreading processes compute over meta-representations. As Moore (2016) points out, success on implicit, 1^{st}-order, false belief tasks is not good evidence for implicit higher-order mindreading skills. It doesn't rule the latter out, of course. But it represents an absence of evidence nonetheless. Scott-Phillips' proposal also rests on findings that higher-order belief reasoning among linguistically competent subjects is relatively quick and effortless (O'Gradey at al. 2015). But as Moore (2016) also points out, this is problematic because linguistic abilities might make higher-order mindreading possible in the first place.

It may turn out that System 1 computes meta-representations of recursive communicative intent. This would depend on further empirical investigation into implicit higher-order mindreading among young children. One study suggesting that the hypothesis deserves to be taken seriously is by Grosse et al. (2010), mentioned in passing in Section 1. The study suggests children as young as 18 months gesture, intending not just that their gesture has a certain effect on another, but also that the hearer *recognise* this intention. That is, the study suggests children as young as 18 months act with communicative intent[75]. Although it didn't test the intentional structure of recognising (as opposed to expressing) communicative intent in young children, the fact that the study provides evidence that children of 18 months intend$_1$ H to recognise$_2$ that they (i.e. the child) wants$_3$ H to do X gives us reason to think that children this young are able to *recognise*$_1$ that S intends$_2$ they (i.e. the child) recognise$_3$ that S is communicating.

The study in question required infants to request an object they wanted from a first experimenter (say, a ball to complete a task the infant was playing with a second experimenter). The first experimenter would then 'misunderstand' the infant's request for the ball by pointing to a different object (e.g. a bit of paper) and uttering 'Oh, you want the paper!'. However, at the same time as reaching for the object the infant did not want, the first experimenter would distractedly place the requested object (the ball) in easy reach of the infant, so that the infant's informative intention was essentially satisfied, despite being misunderstood. This 'misunderstanding' condition was compared with a series of other conditions, the most relevant being the condition

[74] More precisely, 3-year-olds typically fail, 4-year-olds display some success, and 5-year-olds typically pass.

[75] Indeed, with Gricean (i.e. non-Moorean) communicative intent: S intends$_1$ that H recognise$_2$ S intends$_3$ H to do X.

where the infant got the ball *and* was understood correctly. It was found that around 45 per cent of 18-month-old infants attempted to 'repair' the communicative interaction in the misunderstanding condition. Attempts to 'repair' were defined as "all verbal utterances, vocalizations, points, reaches, and showing gestures that were directed at E1 and were not clearly affirmative in nature (as indicated by nodding and/or smiling)" (Grosse et al. p. 1715, my italics). Also measured was the infants' latency to turn away from this communicative exchange and to re-engage with the game the infant was previously playing with the second experimenter. It was found that, when an infant got what it wanted but was misunderstood, the time taken to turn away from the communicative exchange was a bit more than double the length of time (around 9 seconds) compared to when the infant got what it wanted and was understood correctly (around 4 seconds). The conclusion of the study is that infants communicate, not just to have their (informative) intentions met, but also to be understood.

This study can quite easily be interpreted as showing that the infants acted not only with informative intent, but also with communicative intent: i.e. that the infants had the goal, not just that their utterance had a certain effect on the adult, but *also* that the hearer *recognise* that they (the infants) had this goal. That is, it can be taken to show that infants (as speakers) intended$_1$ that H believe$_2$ they (the infant) wanted$_3$ the ball. If this hypothesis is taken seriously, then certain empirical predictions follow. Specifically, we should expect young children to pass implicit higher-order mindreading tasks around the time they begin communicating. If System 1 mindreading abilities allow young children to express recursive communicative intent, then these same implicit abilities should allow subjects to pass implicit mindreading tasks testing for the same level of intentionality.

2.2 SYSTEM 1 OPTION B: IMPLICIT COORDINATION OF ATTENTION

Now for the second version of an implicit solution to the conundrum presented in Section 1. On this version, the processing of communicative intent takes place in System 1. However, the relevant representation in this instance is not isomorphic with the explicit content (of communicative intent): *S intends*$_1$ *that H recognise*$_2$ *S is attempting*$_3$ *to communicate*. There is no higher-order mindreading going on, no meta-representation, even implicitly. Instead, System 1 gives rise to communicative behaviours mimicking those produced on the back of recursive, communicative intent. Communication is thus dual-layered, and hence continuous with Gricean communication, but does not involve the expression and/or recognition of recursive communicative intent. This is a more general view of communication: one which includes Gricean (and post-Gricean) communication within it. The view is 'inferential' when it comes to working out the speaker's intended message (her informative intention), but is purely code-based when it comes to initiating communication (and hence there is no need for the expression and/or recognition of communicative intent).

The view draws from one interpretation of Csibra (2010). On it, communication can be initiated in one of two ways. First, in Gricean fashion: i.e. by the speaker providing good evidence that she is attempting to communicate, from which the hearer must then *infer* that the speaker has an impending informative intention. As a result of inferring the speaker's communicative intention, inferential-interpretive processes (something like those outlined in Sperber & Wilson 1986) are triggered in expectation of further actions produced by the speaker evincing her (the speaker's) informative intention or message. The second way communication can be initiated is by the speaker providing an *ostensive signal*, which non-reflectively triggers the inferential processes in the hearer required to interpret the speaker's informative intention.

The difference between these two ways of initiating communication comes down to the nature of the procedure that triggers inferential-interpretive processes in the hearer (i.e. processes related to representing common ground and interpreting the actions of the speaker in line with something like the principle of relevance). In the first case, the hearer *infers* the existence of communicative intent, before engaging in *further* inferential processing of informative intent. Also, in the first case a meta-representation must be entertained (i.e. a representation of the speaker's communicative intention). In the second case, however, the speaker need only infer informative intent. The inferential-interpretive processes that were inferentially triggered in the first case are, in the second case, automatically trigged. Also, a meta-representation of the speaker's communication intention is not entertained.

'Motherse' and eye-contact are two examples given by Csibra of ostensive signals: codes which automatically trigger infants to "make an effort to fill the empty placeholder of an informative intention" (Csbira 2010, p. 154). In response to such ostensive signals, infants automatically employ certain assumptions that help guide their inferential search for the speaker's informative intention. These assumptions "constrain the search space that infants survey in the effort to attribute an informative intention to the communicator, or in other words, they implement biases that will influence the interpretation of ambiguous communicative acts" (Csibra 2010, p. 154). For example, one of the assumptions triggered (in an infant) by an adult making eye contact with the infant, before fixing on an object and uttering a word, is that this word is being uttered with the intention on the part of the adult to *refer* to the object (and also to treat the object as a member of a general kind)[76].

If something like this proposal is true, then a form of communication exists that is neither code-based nor Gricean, contra Scott-Phillip (2015). Scott-Phillips argues that communication must be either code-based: i.e. involving no real-time/intentional coordination between speakers and hearers; or else 'ostensive-inferential': involving the expression and recognition of recursive intentional states. But on my way of interpreting the Csibra (2010) proposal, the available options are more nuanced than Scott-Phillips suggests. According to this System 1 hypothesis, communication requires the expression and recognition of informative intent, but *not* the expression and recognition of communicative intent. System 1 manages the coordination required to initiate communication successfully, but the relevant System 1 computations manage this coordination without processing meta-representations, the contents of which are isomorphic with explicit descriptions of communicative intentions. The relevant coordination (allowing the search for relevance to be initiated in the hearer, for example) is due to historical coordination over phylogenetic or ontogenetic timescales, as opposed to the real-time coordination of intentional states. For example, Csibra (2010) can be taken to suggest that children are born with the hard-wired tendency to search for an informative intention lying behind an adult's behaviour when this behaviour is preceded by, or accompanied by, ostensive behaviours: behaviours such as making eye contact or talking in motherese. But the child need not represent (in System 1) *that an adult intends them to recognise that communication is taking place.*

Like the last proposal, the current one makes certain empirical predictions. Specifically, children should be able to begin communicating successfully before they show competence in controlled testing of implicit higher-order mindreading. As highlighted in Section 2.1, research is needed to determine whether the shift to implicit testing reduces the age at which children pass higher-order false-belief tests. If it turns out that young children display pragmatic competence before they show competence in implicit higher-order mindreading tests, the present System 1 proposal would be favoured over the first System 1 proposal presented in Section 2.1.

Before summarising, where might Apperly & Butterfill's ('AB') currently-influential 'minimal theory of mind' fit with the current proposals for handling the initiation of communication via System 1? AB (Butterfill & Apperly 2013, Apperly & Butterfill 2009) propose a dual-system account of mindreading, drawing upon considerations from developmental, comparative and cognitive psychology. They

---

[76] It must be noted that Csbira's (2010) proposal is a live research program, and some of the supporting evidence has been challenged: see Gredebäck et al. (2018), Heyes (2016), Moore et al. (2013), and Moore et al. (2015).

propose that the success of infants and nonhuman animals on some belief-reasoning tasks might be explained by an efficient and relatively automatic system for tracking mental representations. This system is innate and runs in parallel to a more flexible, but slow and effortful system, of explicit theory-of-mind developing later in life. They are motivated by the very plausible idea that mindreading abilities are subject to competing demands for efficient and flexible processing. On the one hand, mindreading abilities must be fast enough to facilitate competitive and cooperative social activities in dynamic situations: like trying to figure out what your team mate with the ball will do upon seeing an opponent move out of position to cover you as you lead for the mark. On the other hand, mindreading abilities need to be flexible enough to facilitate explicit reasoning about action in 'offline' situations: like trying to figure out why your partner wrote 'I'm not happy with the way things have been going recently' in the note left on the kitchen bench.

AB's proposal is that humans possess two systems for mindreading: System 1 which is fast, efficient and relatively automatic; System 2 which is slow, flexible and relatively effortful. Moreover, they see the representations involved in System 1 mindreading as being very different in content and functional role from those involved in System 2 mindreading. The former representations are not isomorphic to language-based representations of mental states, while the latter are. Nor is the functional role of the relevant System 1 representations as rich and inter-linked as that of System 2 representations of mental states. The relevant System 1 representations of mental states are much less inferentially integrated with other such representations, which explains why processing over them is more efficient.

Now, one might think AB's System 1 mindreading capacities are a great candidate for the implicit processing of recursive communicative intent (in the vein of section 2.1), thereby making a Gricean account of communication tenable in the face of the relevant ontogenetic constraints presented in Section 1. However, here I cannot appeal to AB's implicit system as a way of saving a Gricean account on this score. For, I argued in Section 1 that the (functionally-distinct) communicative intention must be recursive to do the work required of it, despite the informative intention not being embedded within it. AB's implicit system for mindreading allows the likes of infants, chimpanzees, scrub-jays and humans under cognitive load to track others' representations. But it doesn't allow them to track others' representations of others' representations. This doesn't mean AB's proposal is irrelevant, however. On the contrary, AB's proposed system for implicit mindreading is a perfect candidate for a mechanism handling the fast and efficient processing of *informative intent* when communication occurs 'on the fly'. The proposal presented in Section 2.2 holds that initiating communication could be code-based (Csibra 2010), while the grasping of informative intent (i.e. speaker meaning) is inferential. But it is fairly clear that, in cases of communication occurring 'on the fly', this inferential process must be fast and efficient. Unlike when trying to work out the meaning of your partner's note left on the kitchen bench, grasping the meaning of 'look out!' yelled on the football field is not a complicated process placing heavy demands on executive control and requiring an exhausting search of background knowledge. Instead, you quickly take note of where the speaker is looking and the context, inferring unreflectively that an opponent is probably charging you from your right. This is a cognitive feat well within the bounds of AB's implicit system[77].

Time to take stock. Post-Griceans face a puzzle when it comes to communication. I argued in Section 1 that a deflated picture of post-Gricean communication, involving functionally-distinct communicative and informative intentions, still requires young children to possess recursive mindreading abilities. This is in tension with studies of (explicit) recursive mindreading in children showing that it is only at age 5 or 6 that they possess 3rd-order intentionality (Miller 2009). One solution is to throw out the Gricean picture altogether. A second is to take seriously the idea that System 1 processes enable higher-order mindreading among young children. The last option is to posit a form of communication that lies in-between Gricean communication, on the one hand, and purely code-based communication, on the other. This kind of communication requires the expression and recognition of informative intent, but not necessarily the expression and recognition of communicative intent. Instead, the relevant coordination required to initiate communication could be managed by a code (i.e. natural meaning), while the means required to grasp the speaker's intended message could be managed by inference and meta-representation.

In what follows, I refer to the kind of meaning expressed and recognised in this novel form of communication 'meaningC' (after Csibra). On the one hand, meaningC is contrasted with meaningNN in not requiring meta-representational inference about communicative intent (whether functionally-distinct or not). On the other, meaningC is contrasted with natural meaning in requiring the expression and recognition of informative intent. However, meaningC involves more than just the expression and recognition of informative intent (what I will call 'bi-intentional communication': see Figure 1). With meaningC, but unlike bi-intentional communication, the expression and recognition of informative intent is constrained by the principle of relevance, or something functionally similar (like Grice's maxims). Like with meaningNN, when it comes to meaningC the interpretation processes is aided by cognition taking the principle of relevance (or the maxims) as a constraint on

[77] Heyes (2014, 2015) argues that it is an open question whether the implicit cognitive processes at work in cases like these are functionally dedicated to (an implicit form of) mindreading, as Apperly & Butterfill maintain, or whether they might be handled by (implicit) domain-general processes.

what the speaker intends by uttering. However, unlike with meaningNN, the processing constraints that contribute structure to the inferential search for the speaker's meaningC need not be triggered by a meta-representation of communicative intent (see Figure 5.1 below).

Having charted the options just outlined, I now consider where nonhuman primates fit in: specifically chimpanzees and old-world monkeys. Moore (2016) argues that chimpanzees communicate in the gestural modality motivated by both informative and communicative intent, placing them within the realm of meaningNN (Figure 5.1). Cheney and Seyfarth (2008) (also Seyfarth & Cheney 2017) argue that old-world monkeys express and recognise informative intent in the vocal modality, effectively placing them within the realm of bi-intentional communication (Figure 5.1). I am sceptical of both appraisals. Explaining why I am sceptical will also help illuminate the options presented in Figure 5.1.

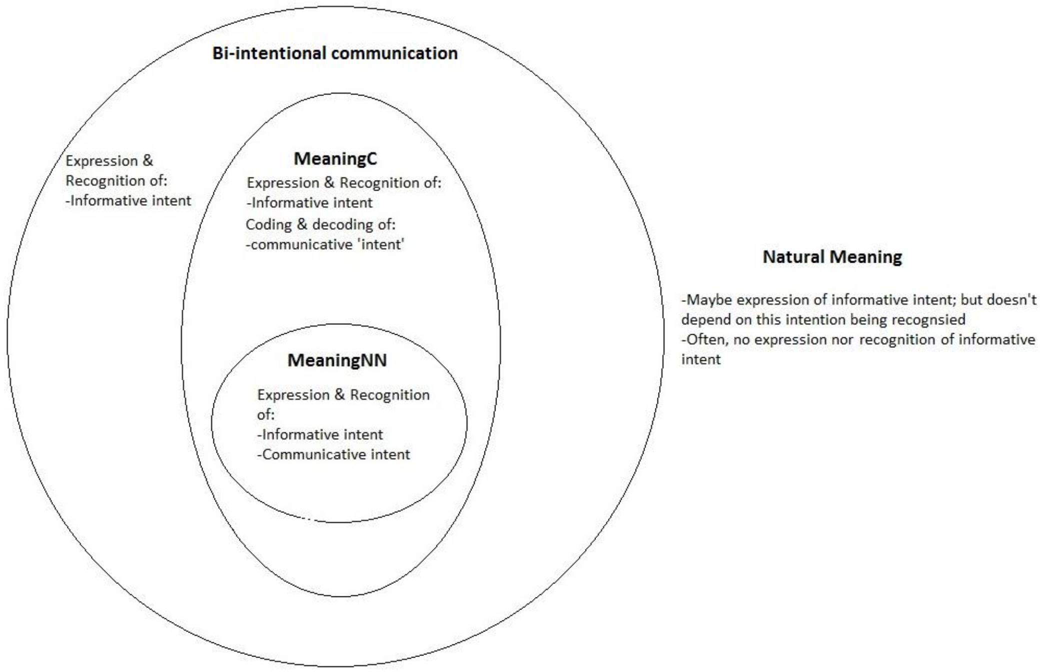

Fig 5.1 Grades of meaning, from least to most sophisticated (outside-in). Natural meaning needn't be the product of intentional behaviour on the part of senders. Sometimes, however, natural meaning can be sent intentionally, as when I purposefully accentuate to you that I have spots by giving you a clear view of them. However, the recognition of natural meaning requires no meta-representation on the part of receivers. Bi-intentional communication requires the expression and recognition of informative intent. MeaningC also requires this, in addition to signalhood being indicated and recognised (but the latter, without meta-representation). MeaningNN requires, in addition to the expression and recognition of informative intent, the expression and recognition of communicative intent – whether functionally distinct or not.

3. Ostension and inference in nonhuman primate communication

- Nonhuman great apes and ostension –

Moore (2016) argues that when chimpanzees gesture, they express both informative and, more controversially, communicative intent. Chimp gestural communication is dual-layered in the Gricean sense. In Tomasello's (2008, p. 22) comprehensive review, two basic types of (intentional) great ape gestures are described: intention-movements and attention-getters. Intention-movements are abbreviations of social actions (like playing or pulling mum's back down to climb on it), where receivers have come to anticipate the sender's behaviour over repeated instances of the interaction. For example, a young chimp approaches another wanting to play, raises his arm in preparation to play-hit the other, and then goes ahead and hits, which initiates playing. Over repeated instances of this, the receiver comes to learn to anticipate the sequence on the basis of the initial arm-raise, and so begins playing after perceiving just this first step. The sender, respectively, also come to anticipate the receiver's behaviour, and so simply raises his arm expecting the receiver to initiate play.

Attention-getters are different. They function to draw the receiver's attention to something the sender thinks will cause the receiver to do what he (i.e. the sender) want. For example, male chimps wanting sex sometimes engage in 'leaf-clipping' behaviour: making sharp, loud noises (presumably with leaves) and attracting the attention to females to their erect penis. Attention-getters are presumably learned by individuals doing something for non-communicative reasons which naturally attracts the attention of others, a result which is noticed and then exploited in the future in the service of informative intent (Tomasello 2008). As an extension of their natural attention-getting gestures, chimps raised in captivity 'point' for humans who (unlike other chimps) are disposed to assist chimps in getting them what they want. Chimps point to out-of-reach food they want; to tools a passing human could use to retrieve the food for them (but not for tools the human could use for his own benefit); and to locked doors they want access behind. Chimps in zoos also clap their hands, or engage in other attention-getting behaviours, for human visitors so that they will be attended to and, hopefully, thrown food (Tomasello 2008).

Attention-getting gestures like this would seem to show that chimps understand others have perceptions (loosely defined), and that they (i.e. chimps) need to manipulate these perceptions in the service of getting their informative intentions satisfied. Further, Leibal et al. (2004) found that chimpanzees (along with bonobos and gorillas) move into the line of sight of a potential human food donor before gesturing for food. In other words, the understanding chimps have of others' perceptions is intentional, in the sense that chimps *flexibly* seek to gain the attention of conspecifics. As a result, chimps' attention-getting gestures seem to be motivated by something quite like Gricean communicative intent. Recall me making eye-contact and raising my eyebrow to Rick before leaning back on my chair. This seems to be a behaviour with a very similar function to a chimp's efforts at directing the attention of a receiver by pointing, clapping, moving into their line of sight, or making loud noises for them with a leaf.

The issue I want to raise concerns the attention-getting behaviours of chimps: more specifically, whether such behaviours are motivated by communicative intent[78]. Moore (2016, pp. 226-227; 2017, p. 316) argues that if S performs one action (say, moving into H's line of sight) as a means of addressing a second action (say, a begging gesture) to H, then it is true to say that S intended$_1$ H to recognise$_2$ that S wanted$_3$ food (perhaps implicitly). I do not see a good reason to grant this conditional claim. This is because informative intent alone might have motivated the very same action on the part of S. It is just as plausible that S moves into H's line of sight, not because S intended$_1$ H to recognise$_2$ S wanted$_3$ food, but rather because S knew$_1$ H must see$_2$ her begging gesture for it to elicit food from the human. There need be no (explicit nor implicit) representation on the part of S about the perspective taken by H towards her (i.e. S's) intentions. All there need be is a representation of the part of S about the perspective taken by H towards the outward gesture: '*Can H see$_1$ my begging gesture?*'[79].

S's attention-getting behaviours might be performed because of an instrumental belief underlying informative intent alone, as opposed to anything like communicative intent (whether represented implicitly or explicitly). After all, if informative intentions are *intentions*, i.e. mental states that produce behaviour robustly sensitive to means-ends contingencies, then we should expect an animal with an informative intention (i.e. that a human produce food) to act in flexible ways to have this intention fulfilled. This, combined with findings that chimpanzees understand what others can and can't perceive (Hare, Call & Tomasello 2001), gives us no reason to draw a necessary equivalence between attention-getting behaviours, on the one hand, and behaviours motivated by communicative intent over and above informative intent, on the other. Of course, sometimes attention-getting might be done with something like communicative intent. But there is no necessary equivalence between getting the attention of an intended receiver, on the one hand, and communicative intent, on the other.

Does this mean chimps are restricted to naturally-meaningful communication, like fireflies? No. Chimps are not just coding and decoding signals. Senders act intentionally to bring about some effect in the hearer, and, crucially, hearers may be inferring something like the sender's intention in gesturing. If this is true, then chimps

[78] I argued earlier that the functionally independent communicative intention carries one more order of intentionality than Moore suggests, but this does not affect the current point.

[79] An even more deflationary hypothesis that would account just as well for chimpanzee acts of address is that S simply knows$_1$ that she must be in front of H before begging. I.e. there need not even be an awareness of what others see. However, studies suggest that chimps represent, at least in some way (perhaps akin to our System 1 abilities in this domain), what conspecifics are attending to (Hare, Call & Tomasello 2001). So, while it is possible that chimps are not bringing this capacity to bear on communication, it seems unlikely.

are, at the very least, bi-intentional communicators (Figure 5.1). If indeed chimp gestural communication is bi-intentional, then this still sets it apart from most of the signalling systems considered in this thesis (including old world monkey vocal communication, as will be argued shortly).

Might chimp gestural communication traffic in meaningC? This would depend on whether communication is 'cooperative', in the following sense. Are speakers choosing behaviours, to get what they want via communication, constrained by how hearers interpret? Are hearers making inferences about what speakers intend, guided by the expectation that speakers are so constrained? If not, then chimps won't be communicating meaningC. They will be bi-intentional communicators, only. Even if chimps are not communicating meaningNN nor meaningC, however, the fact that chimps understand they need to address their communicative actions to conspecifics is plausibly significant. It might create the space within which meaningC and meaningNN develops phylogenetically. While chimp communication is not dual-layered when it comes to its intentional basis, because it often manifests in two separate behaviours (attention-getting/directing and then a signal), there might be a solid basis here upon which meaningC or meaningNN developed.

- Old-wold monkeys and inference -

A recent paper called *the origins of meaning in animal signals* pushes continuity between primate communication in the vocal domain and human linguistic communication. In it, Seyfarth and Cheney (2017) seek to highlight 'pragmatic' competencies in the vocal comprehension abilities of old-world monkeys. One of their main case studies is the baboon. Baboons have rich and busy social lives, living in groups of 50-150 individuals. Males emigrate to other groups as young adults, while females remain within their natal groups for life where they keep close social bonds with their matrilineal kin. Females are ranked in linear dominancy hierarchies that determine access priority to resources. Daughters acquire a status like that of their mothers. All members of one matriline (for instance line B) outrank or are outranked by all members of another matriline (i.e. C and A, respectively). The result is a set of relations like the following: A1>A2>A3>B1>B2>B3>C1>C2>C3, where letters denote matrilineal kin groups and numbers denote individuals within each kin group (Cheney & Seyfarth 2007).

Over the course of any single day, baboons hear a range of vocalization. Threat-grunts are an exemplar. Here, a higher-ranking individual grunts to a lower-ranking individual. In response, the latter usually screams: a submissive signal given primarily by lower to higher-ranking individuals. Another vocalisation is the reconciliatory grunt, given after conflicts to minimise the disruptive effects of within-group tension. Reconciliation occurs after about 10% of all fights, and occurs when the dominant animal grunts to the subordinate. The typical result is that the victim will tolerate the approach of her former opponent, or alternatively approach her former opponent herself. Importantly, vocalisations are individually distinctive: receivers recognise the identity of senders from voice alone. Moreover, calls are given in predictable social contexts. As a result, baboon receivers can acquire quite specific information from combining a representation of who called with a representation about the social context in which they called. For instance, when a baboon hears a member of line C produce a threat-grunt which is followed by a scream from a member of line B, she displays much more interest than if she hears a member of line C scream in response to the threat-grunt of a line B member. This is presumably because, in the first case, she has inferred a between-family rank reversal taking place. Within-family rank reversals also elicit surprise, but baboons show more surprise in response to apparent between-family reversals. Most pertinent to the current discussion are observations, which, according to Seyfarth & Cheney (2011)[80], show that baboons recognise "other individuals' intentions and motivations". If baboons do indeed recognise sender intent during communication, then this places them in the category of bi-intentional communicators.

For example, when a female hears the threat-grunt of a recent opponent, she responds as though the grunt was "directed at" her. However, when she hears the same female's threat-grunts soon after grooming with her, she acts as though the grunt was "directed at" a different baboon. Another example concerns reconciliatory grunts. As a proxy for direct reconciliation by a former opponent, baboon victims often accept the reconciliatory grunt of a close relative of the recent opponent. For example, when A2 grunts in the hour following a confrontation between C1 and A1, then the former victim (i.e. C1) is more likely to tolerate the approach of A1 or A2 compared with: a) if C1 heard no grunt at all; or b) if C1 heard a grunt from another higher-ranking individual unrelated to the A matriline. From observations like these, Seyfarth & Cheney allege that baboons supplement the meagre information available from a signal[81] with contextual information. As a result of this process of "pragmatic inference", baboons infer whether the sender's intention in grunting concerns them or another baboon. They draw the conclusion that "in the primate lineage, pragmatic inference seems likely to have served as a precursor to the evolution of semantics, syntax and language among humans" (Seyfarth & Cheney 2017, p. 8).

This conclusion is outwardly like Scott-Phillips (2015), where ostensive-inferential communication, or a 'pragmatics-first' picture of communication, is alleged to make the advent of syntax and thus human language possible. However, what is

---

[80] See Cheney & Seyfarth (2007), where they also claim intention recognition is taking place.
[81] Close to the sense of natural meaning I defined in Chapter 1.

meant by the label 'pragmatics' differs significantly between Scott-Phillips, on the one hand, and Seyfarth & Cheney, on the other. For both parties, pragmatic interpretation involves inference on the part of receivers: that is, some kind of rational process operating over representations of information in order to reach a conclusion about the significance of a signal. However, in the pragmatics-first picture defended by Scott-Phillips, the inferential process results in a representation of what the speaker intended by producing an utterance. When it comes to baboon pragmatics, on the other hand, the conclusion that the inferential process results in a representation of sender intent is questionable.

Consider reconciliation grunts. There need be no inference, on the part of listeners, regarding who A2 *intended* to reconcile with her relative, A1. C1 only need infer what state of the world has been made more likely, given A2's grunt and also C1's recent conflict with A1: namely, that approaching A1 (or A2) will not be dangerous now. Consider also threat grunts. Recall that when a female hears the grunt of a recent opponent, she responds as though the grunt was 'directed at' her. However, when she hears the same female's threat-grunts soon after grooming with her, she acts as though the grunt was 'directed at' a different baboon. Some kind of inferential process has obviously taken place here, but the conclusion of this process needn't be a representation about sender intent. Instead, the conclusion could refer simply to the likely state of the world, given the grunt and the listener's recent dealings with the sender. If the listener has recently fought with the sender, then behavioural evidence that she (i.e. the listener) has taken the threat grunt to be 'directed at' her does not imply meta-representation of sender intent. 'Directed at', in the mind of the listener, could just amount to 'has ramifications for me as opposed to another baboon'.

The issue is unlikely to be settled by playback experiments alone. The observed behavioural responses of listeners are compatible with either inference about sender intentions or merely inference about the state of the world. In order to establish which hypothesis is more likely, it might be helpful to turn to the mindreading abilities of monkeys more generally. A recent review (Meunier 2016) splits mindreading into 5 components and presents the state of play regarding each: gaze following, perspective-taking, attention reading, intention comprehension, and false-belief understanding. When it comes to gaze following, monkeys seem to resemble apes in following another's gaze, particularly when the other individual is a conspecific. The evidence for (level 1) perspective taking in monkeys is slightly more "mixed". As with chimps (e.g. Hare, Call & Tomasello 2001), there is some evidence that long-tailed macaques, ring-tailed lemurs, and rhesus macaques can represent what conspecifics are perceiving in competitive feeding paradigms, but only under certain experimental setups. When it comes to attention reading, monkeys can discriminate gross cues like body and face orientation while chimps, in addition, possibly display more sensitivity to the state of the eyes. The most relevant component, intention comprehension, is characterised by contradicting results and "new data are eagerly awaited". As of the time of the review, no monkey species has been shown to represent the false beliefs of others.

Thus, it seems reasonable to conclude that more information is needed before classifying old-world monkeys like baboons as bi-intentional communicators (Figure 1). There is currently no good reason to conclude that the following conjunction is require for successful communication: a) that senders act with informative intent towards receivers; *and* b) that receives need to infer the sender's informative intention. The first conjunct seems likely, while the second is an open question. The sole fact that baboons engage in perhaps relatively sophisticated, domain-general, inference to determine the significance of vocal signals does not show that baboons infer informative intent. At most, receivers might be just representing a relatively wide range of factors in inferring the *natural* meaning of the signal.

4. Conclusion

This chapter focused on two main issues. First, ways of deflating the intentional requirements of post-Gricean communication in response to ontogenetic constraints. In Section 2, I argued that those wanting to hold on to meaningNN must take seriously the hypothesis that humans are born with the ability to meta-represent using System 1. I also made space for forms of intentional communication not involving the expression and recognition of communicative intent. Second, Section 3 argued that there is no good reason to infer from chimpanzees' acts of address that they meanNN things when gesturing. Instead, chimps may be bi-intentional communicators. Despite this, the fact that chimps might well be attention-readers makes their communication an interesting precursor to human communication. I also argued that the pragmatic-inferential abilities of old-world monkeys do not yet show them to be bi-intentional communicators, despite these abilities plausibly representing an important step in the road towards the domain-general nature of human comprehension.

*   *   *

Chapter 6. *The expression and recognition of intentional states is a crucial component in explaining how human beings express and recognise speaker meaning. But what is the explanatory status of intentional states? Can intentional states feature in genuine cognitive explanations of some capacity, like communication? What is the relationship between intentional explanation and cognitive explanation?*

Chapter 6

The Explanatory Status of Intentional Psychology

The main posits in post-Gricean explanations of communication are intentional states. Chapters 4 and 5 referred to 'communicative' and 'informative' intentions, although some have different labels for these intentions. However, the main issue is the structure and role of these intentions in successful communication, and also the consequences of this structure for the meta-psychological requirements of communication. A range of researchers agree on one fundamental point: that Grice's analysis of meaningNN revealed some of the key psychological capacities necessary for communication (Moore 2017; 2016, Scott-Phillips 2015, Tomasello 2008, Sperber & Wilson 1995; 2002). The capacities identified by Grice's analysis explain pragmatic aspects of communication, and anything less than this intentional structure fails to produce the kind of communication humans engage in: i.e. the communication of meaningNN as opposed to natural meaning. What some researchers disagree on are the finer details of this intentional structure: whether communicative and informative intent must be entwined in a single intentional state, i.e. where the content of the latter is embedded in the former (Scott-Phillips 2015, Sperber & Wilson 2002); or whether communicative and informative intent can be functionally distinct (Moore 2017). This difference is important because of its ramifications for the meta-psychological demands of communicating meaningNN. If communicative intent can be functionally distinct from informative intent, then somewhat less recursive mindreading is required by interlocutors. This, in turn, is important because of its ontogenetic and phylogenetic implications (as outlined in the previous chapter).

In Chapter 5, I examined the meta-psychological demands of communicating meaningNN, assuming communicative intent and informative intent *are* functionally distinct as suggested by Moore (2017). However, I parted ways with Moore on the structure of the communicative intention. I did this because of the role of the communicative intention in signalling signalhood, and thus relevance (or something functionally similar). Here, again, is the intentional structure I proposed:

1. Are signals produced by senders <u>intending</u> to elicit a particular response from a receiver? If so, does the receiver <u>recognise</u> the sender's intention?
    - This criterion concerns the informative intention
2. Do senders <u>intend</u> their audience to <u>recognise</u> that they are communicating? If so, does the receiver <u>recognise</u> that the sender <u>intends</u> them to <u>recognise</u> this?
    - This criterion concerns the functionally-independent communicative intention

As discussed, if these abilities are indeed required for communicating meaningNN then it doesn't sit comfortably with the idea that children can communicate successfully before they succeed in recursive mindreading. One option is to give up altogether on Grice as *explaining* communication. If meaningNN is so closely tethered to an intentional structure beyond the capabilities of young children who communicate, then in giving up this intentional structure we are giving up on meaningNN. The transition from natural meaning to meaningNN will *not* be the transition from nonhuman to human communication[82]. Perhaps, as suggested in Section 1 of Chapter 4, the post-Gricean approach took an analysis that was not intended to be explanatory in the wrong direction.

In response, I explored a less pessimistic route. Even if meaningNN *is* strictly speaking outside of the capabilities of young children (and also our closest living ancestors), I argued that human communication, at its simplest, may involve the expression and recognition of informative intent alone: where this process may be assisted by the activation of certain communicative assumptions using ostensive signals (Csibra 2010). For instance, simply hearing you utter an English sentence might prime me to search for the relevance of your linguistic action, making it easier for me to infer your informative intention from your outward behaviour. But there is no reason to demand that my response is a product of meta-representing communicative intent. Likewise, a child being addressed in motherese might attempt to fill the placeholder of an informative intention. But this need not be the result of the child <u>recognising</u>$_1$ that S <u>intends</u>$_2$ H to <u>recognise</u>$_3$ that communication is occurring. Instead, motherese just needs to prime H to search for informative intent, via a code (i.e. a naturally-meaningful signal). This represents a form of communication standing midway between purely coded communication, on the one hand, and Gricean non-natural meaning, on the other. I also presented a second form of communication that is neither purely code-based nor meaningfulNN. It involves the expression and recognition of informative intent, without ostensive signals. I called it 'bi-intentional communication'. On the options presented in Chapter 5, then, there are grades of

[82] Nor will it represent the transition from code-based communication to something resembling human communication, if one thinks this transition occurred earlier.

speaker meaning. There is the kind of speaker meaning plausibly expressed and recognised by chimps in their bi-intentional communication. Then there are kinds of meaning humans express and recognise, if indeed communication is 'cooperative' in the sense that speakers and hearers are (perhaps implicitly) constrained by something like the principle of relevance: meaningC and meaningNN.

1. Accounts of intentional mental states

I now want to change tact and raise a more fundamental issue. The issue is the explanatory status of intentionality. This issue has major implications for the kinds of speaker meaning just discussed. All rely heavily on the theoretical posit of intentionality. However, there are mutually-inconsistent ways of understanding intentionality in the philosophical literature. Some philosophers hold a very permissive view of intentionality. Alternatively, some hold a very demanding view of intentionality ruling out non-linguistic organisms as having intentional states. Put simply: what *is* intentionality (or, what I take to be equivalent: what *is* an intentional mental state)? Intentional states are commonly taken to be belief-desire pairs, and beliefs and desires are folk states. Crucially, how folk-psychological explanations of behaviour involving beliefs and desires interface with cognitive explanations of behaviour is not straightforward. In fact, it is perhaps the most fundamental issue in the philosophy of psychology (Bermudez 2005).

One kind of response to the interface problem takes folk-psychological states to correspond quite directly to cognitive states, such that folk states are genuinely explanatory of behaviour. This view posits a significant isomorphism or 'image' of the propositional attitudes in the structure of cognition. It proposes a type-type reduction of folk-psychological states to cognitive states (but not typically neural states). Fodor's representational theory of mind (RTM) is a paradigm example of such a view. According to Fodor's RTM, the states that drive thought and behaviour from the perspective of the cognitive sciences are structurally isomorphic, on a suitable level of abstraction, to the linguistic descriptions we give of beliefs and desires. This is why Fodor's RTM is called the 'language of thought'.

On the other end of the spectrum lies Dennett's response to the interface problem[83]. Dennett is prepared to attribute beliefs and desires to any system in which doing so allows us to predict its behaviour more powerfully than we could from the functional and physical stances alone. If we benefit from considering what it would be *rational* for the system to do given its function(s) plus its source(s) of information about the world, then the system in question has beliefs and desires. However, in

[83] Davidson and McDowell's responses to the interface problem also lie on this end.

classifying some system as a belief-desire/intentional system, Dennett is not attributing discrete states driving thought and behaviour and corresponding to linguistic representations of belief and desires. There is no type-type reduction of folk psychological states to cognitive states. Instead, the relevant beliefs and desires are "emergent" upon a complex conglomerate of cognitive states that might bear no correspondence to a belief-desire explanation of behaviour. Think of how a bird flies without internally representing aerodynamics. Similarly, according to Dennett, agents act rationally (or semi-rationally) without internally representing beliefs and desires. In between a Fodorian response to the interface problem on the one hand and a Dennettian response on the other lie various forms of functionalism.

Somewhat congruent with these different responses to the interface problem are different views on the ontology of intentionality. We will start with the most minimal account of intentionality. Dennett holds that an organism is an intentional agent just in case the strategy of taking that organism to be an agent is needed to predict its behaviour. This is taking the *intentional stance*. Taking the organism to be an agent, or assuming the intentional stance towards the organism, involves attempting to predict its behaviour by attributing to it proposition attitudes (beliefs and desires). But by doing so one is *not* attributing (discrete) internal states to the system. Instead, one is simply attributing course-grained aptitudes for world-directed behavioural patterns (Dennett 1992, Matthews 2007). More precisely, one is not making a claim about internal cognitive states beyond committing to the idea that there is *something* going on giving rise to the relevant behavioural patterns[84]. What are these behavioural patterns? As mentioned in Chapter 3, they involve the ability to pursue the same end via different means, and different ends via the same means. In short: behavioural resilience. An intentional system can act (fairly) adaptively if the environment is tweaked somewhat. When a system can do this, according to Dennett, then taking the intentional stance over and above the physical and design stances becomes necessary. Essentially, attributing intentionality is a claim about a specific kind of capacity, as opposed to how that capacity is mechanically supported.

Next up is the most demanding view of intentionality: what I will label The Higher-Order Deliberation View (HOD). According to HOD, only organisms meeting two criteria perform intentional actions (Arruda & Povinelli 2016). The first is the one outlined immediately above: the ability to pursue the same end via different means, and different ends via the same means. Second, the view requires higher-order deliberation: the ability to explicitly represent one's goals *as* one's goals to evaluate,

[84] It is important to note that one can perfectly well be a representationalist about subpersonal cognitive states while taking the intentional stance to be the appropriate attitude towards folk psychological states like beliefs and desires: e.g. Cummins 1989. In other words, while beliefs and desires might not be syntactically structure mental representations, the complex behavioural patterns that warrant taking the intentional stance might be caused by syntactically structured presentations at the subpersonal level.

rank, or otherwise reason about them. An important qualification to the last requirement is needed, however, to make this view of intentionality credible. Much human behaviour occurs in the absence of higher-order deliberation about what one is doing. But this doesn't mean such behaviour is on par with non-intentional behaviour like reflexive behaviours (Brownstein 2014). Take the highly skilled movements of a professional pianist's hands over the keyboard. Such movements are intentional, even though, while the movements were executed, there was no higher-order deliberation going on. Why are the pianist's movements intentional? Because he could probably give reasons for his movements, after the fact. Even if he was not explicitly reasoning about his movements while making them, he could explain why he made those movements to anyone who asked after the concert[85]. One plausible moral to take from this relates to our second criteria for intentional behaviour. Instead of requiring *actual* deliberation, it should instead require actual or *counterfactual* deliberation. On this view, a behaviour is intentional if the individual did, or could have, engaged in higher-order deliberation about that behaviour (Brownstein 2014).

The last account of intentionality to be considered is based on a Fodorian RTM: i.e. a representation theory of mind that is representational in virtue of a language of thought (LoT). Some versions of RTM invoke maps as their representational entities, as opposed to sentences of Mentalese (Braddon-Mitchell & Jackson 1996, Camp 2007, Rescorla 2009). This kind of RTM will not be addressed here, though, partially for reasons of space and partially because they are worked out in less detail. Thus, when I refer to RTM in what follows, I mean an RTM where the representational vehicles are linguistically structured.

Intentionally, on RTM, does not require higher-order deliberation: neither actual nor counterfactual. It is less demanding. But it is more demanding than Dennett's intentional stance. Saying a system has an intentional state (i.e. a belief or a desire), on RTM, is saying that it possesses an internal representation, the vehicle of which is structurally isomorphic with the content of an applicable propositional attitude report of the belief/desire[86]. An example of a propositional attitude report is the statement *Smith believes Pink Floyd were better than The Beatles*. Here, belief is the attitude reportedly taken by Smith towards the proposition that Pink Floyd were better than The Beatles. Another propositional attitude might be *Smith wishes Pink Floyd were better than The Beatles*. Here, desire is the relevant attitude reportedly taken by Smith towards the same proposition: the proposition that Pink Floyd were better than

[85] For instance: 'I shifted my posture after the exposition section in preparation for the rapid trill beginning high up the treble range in the development section'.

[86] When one appreciates the richness of human language, the idea that the isomorphism obtains between, say, English propositional attitudes reports and representations in LoT becomes questionable. It is plausible that Fodor had in mind something like Chomsky's distinction between surface structure and deep structure, where the isomorphism grounding intentional mental states obtains between the deep structure of English (or Chinese) propositional attitude reports and representations in LoT.

The Beatles. This makes sense of the intuition that agents can take different attitudes towards the same proposition. Moreover, agents can take the same attitude towards different propositions, like when Smith believes the above proposition in addition to believing The Beatles were nevertheless a great band. Now, if Smith indeed believes the proposition that Pink Floyd… etc, on RTM, then he possesses an internal mental representation, the structure of which is isomorphic to the content *Pink Floyd were better than The Beatles*. Moreover, this mental representation must play the kind of functional role characteristic of belief. Alternatively, if Smith desires the same proposition, then he possesses an internal mental representation, the structure of which is isomorphic to the content *Pink Floyd were better than The Beatles*, where this representation carries out the functional role characteristic of desire. In short: when someone believes that Pink Floyd were better than The Beatles, the content of that belief is realised by an internal structure that is the vehicle of it's content in the same way as the meaning of the sentence 'Pink Floyd were better than The Beatles' is realized by the structure of the inscription (Bermudez 2005).

Crucially, the organism need not explicitly represent the content of its belief/desire for it to be an intentional agent. The belief or desire could be tokened 'implicitly'. What is meant by the 'implicit' tokening of a belief or desire, on RTM? One plausible answer is to attribute 'implicitly' held belief to anyone possessing a discrete, syntactically structured representation serving as the vehicle of the belief, but where the agent is not engaged in higher-order deliberation about this belief. Further, the representation serving as the vehicle of the belief must possess the functional role of belief: it can't be a representation isolated in early visual processing, for example. A second plausible answer is to attribute the 'implicitly' held belief that *p* to anyone who hasn't tokened an applicable, syntactically-structured, mental representation, but who could easily infer *p* from those syntactically-structured representations he or she *has* tokened. For instance, folk psychological intuition tells us that that, plausibly, Donald Trump believes the earth is larger than an average pickle. However, Trump probably doesn't possess, at all times, a syntactically-structured representation with this content, playing the functional role of belief. Instead, it is relatively easy for Trump to infer such a belief from those syntactically-structured representations he *is* tokening, at any one time. Contrast this to the belief that Bertrand Russell but not David Lewis thought propositions are structured. Trump surely doesn't implicitly believe *this*. On RTM, this is because a) Trump isn't tokening a syntactically-structured representation playing the belief role with this content; and because b) Trump cannot infer this belief from those syntactically-structured representations that he is tokening, at any one time.

Despite their differences, each of the three broad accounts of intentionality just mentioned endorse the following idea. The idea is that intentional behaviour contrasts with the kinds of reflexive behaviours described by Tinbergen as hard-wired 'releasers'. It also contrasts with behaviour learned via simple forms of stimulus-response conditioning and classical conditioning, at least on 'non-cognitive' accounts of such learning processes (Holland 2008). Hard-wired responses to stimuli acquired either by natural selection or via simple, non-cognitive, forms of learning are commonly seen as non-responsive to how the organism represents the connection between means and ends: i.e. rational or proto-rational representational abilities.

This conceptual connection between intentionality on the one hand and rationality on the other can be seen in Heyes and Dickinson's (1995) critique of the idea that cognitive maps are genuine intentional states in the psychology of an organism:

> How would we ever know whether… the "cognitive map" of a bee has intentional properties (for the… bee, rather than the human observer) unless it can control behaviour that is rational with respect to the content of these states? It is not sufficient to appeal to the adaptiveness of the behaviour because the rationality that matters with respect to intentionality is that of the psychological processes of the individual agent, not of the evolutionary process.

From this we also get a distinction between rationality derived from evolutionary processes and rationality derived from the individual agent itself, in real-time (see also Allen & Bekoff 1997). Dennett, too, endorses this distinction and characterises it as one between 'free-floating' rationality on the one hand, and rationality for the agent on the other. I think it is safe to say, that for the three views of intentionality discussed above, the 'rationality' required of intentional behaviour must be rationality for the agent itself. The three views simply tell different stories about what gives rise to rationality for the agent itself. On the modest side, Dennettians who simply require flexible, goal-orientated behaviour are relatively non-committal on what gives rise to rationality for the agent. The kind of representations could be highly distributed, or they could be discrete and syntactic. On the least modest side are those who require higher-order deliberation. Occupying the middle ground is a broadly Fodorian position: agent-level rationality derives from discrete beliefs and desires guiding behaviour appropriately, even if implicitly.

In what follows, I consider the explanatory status of notions of speaker meaning assuming RTM, on the one hand, and then the intentional stance, on the other. I leave behind the HOD view of intentionality. This is because, if intentionality requires higher-order deliberation, then the intentional demands of communication are automatically increased. If, in order to harbour the informative intention that *Rick look over there*, I must be able to meta-represent my goal, then I must not only <u>intend</u> that Rick look over there: I must also <u>consider</u> (or be able to consider) that I <u>desire</u> Rick to look over there. Basically, any problems about the intentional requirements of broadly Gricean accounts of communication will be multiplied: by one order of intentionality, to be precise.

Another potential issue with the HOD view is that it seems to require intentional agents to possess concepts about representations: concepts like *desire*. If this is the case, then individuals without such concepts will not harbour genuine intentional states, and hence won't be genuine communicators, on broadly Gricean views. Perhaps it might be suggested that individuals represent their own mental states in the absence of concepts of these mental states. Apperly & Butterfill's 'minimal theory of mind' (mToM) comes to mind, here. However, mToM is explicitly not a cognitive system that represents mental states as such. Rather, it represents proxy states that track, in the right circumstances, mental states like *believing*, *perceiving* and so on. The states it represents are labelled 'registrations' and 'encounterings' and are not mental states as such. Moreover, Apperly & Butterfill nowhere suggest that mToM tracks the subject's own mental states. There are no representations of *one's own* registrations or encounterings, generated in mToM. You register, but you do not register that you register. Rather, mToM is a system for the circumscribed tracking of others' mental states based on cues like gaze direction. Even if mToM could represent proxies for one's own mental states, though, it is questionable whether this would enable the kind of higher-order deliberation about one's own goals required by a HOD view of intentionality. It is one thing to represent a proxy state that tracks a mental state in certain bounded circumstances. It is another to deliberate about this proxy state for the purpose of evaluation, raking, or otherwise reasoning about it. Arguably, the latter requires concepts of mental states as such. Registrations and encounterings are not propositional attitudes, because (among other things) they do not possess the rich inferential and causal connections with inputs, other mental states, and outputs characteristic of beliefs and desires (Apperly & Butterfill 2009). It is because registrations and encounterings are not represented as having such rich causal relations that processing over them is fast and efficient. Thus, because these proxy states don't connect rationally with inputs, other mental states, and outputs, it is hard to conceive of a process of deliberation over such states that is in the spirit of a higher-order deliberation view of intentionality.

A yet further issue with HOD is that it may require intentional agents to possess language. There is some debate about whether higher-order thought requires language (see, for instance, Botterill & Carruthers 1999). One of the reasons for thinking that it does includes the above-mentioned consideration. That is, it is difficult to conceive of what higher-order thought might be in the absence of concepts of mental states as such. And one natural position to take is that concepts of mental states as such come to us through language. Before we are enculturated into language, we might have representations of proxies for mental states, but such representations will not be of states that get deliberated about. There are other reasons that have led philosophers of psychology (e.g. Dennett 1991) to propose that higher-order thought requires language. They will lead me too far afield here. Suffice it to say that if HOD requires language, then it will not be applicable to theories of communication evoking intentional states in explaining communication.

2. Trying out RTM

Psychological notions of speaker meaning are hostage to the interface problem. How do they fair under RTM? There is a straightforward sense in which these kinds of meaning will be genuinely explanatory if the relation between folk-psychological explanation and cognitive explanation is one of type-type reduction. Assume, plausibly, that a model of some target system is explanatory to the degree it decomposes a system-level function into component functions. If, as RTM states, intentional states correspond nicely to internal components of a cognitive system, then accounts of communication leaning heavily on intentional states will be explanatory. Of course, the issue of abstraction looms large here. Obviously, beliefs and desires are not found at the neural level on RTM. Even still, on RTM the propositional attitudes are not mere patterns of behaviour, as they are on the intentional stance. Rather, they are discrete states internal to the cognitive system at a suitable level of abstraction, in the same way data structures are internal states of a computer at a suitable level of abstraction.

However, there is a potential worry one might have here, given the view of RTM held by Fodor. For Fodor, RTM is a view about central cognition. RTM finds a home for folk psychology in the scientific image of the mind by positing a type-type reduction of propositional attitudes to states of central cognition. Input systems, like vision or language, might themselves operate over linguistically structured representations (modular LoTs), but propositional attitudes like beliefs and desires find their 'image' in the LoT of central cognition, as opposed to the LoTs of modular input systems (Fodor 1983, Bermudez 2005). This makes sense from the perspective of folk psychology. From this perspective, beliefs and desires are states bearing rational relations with each other and with a range of other intentional states. When Jones learns of the first time that his father was, in fact, a spy, this new belief has ramifications for a wide range of his beliefs and desires. Compare this new belief to Jones's mental representation of an edge located in his early visual system. Unlike the new belief about his father which sends shock waves throughout his whole mental life, a representation of an edge in the visual system has very little, if any, cognitive ramifications.

Now, If the propositional attitudes are restricted to central cognition, then we are obliged to posit that much of the work of speaker meaning expression and recognition takes place in central cognition. This presents a potential problem. If, as dual-system theorists commonly conceive of things, all central processing is identified with System 2, then the expression and recognition of intentions will inherit the properties commonly ascribed to System 2 processing. It will be conscious, controlled, effortful, slow, reflective, and demanding of cognitive capacity (Evans 2008, Stanovich 2004). However, it is unlikely that everyday episodes of communication involve the expression and recognition of informative *and* communicative intent in a way that is conscious, effortful, and reflective (etc.). Likewise, the communication of young children. What about intermediary forms of meaning: those lying in between code-based and Gricean communication, requiring the expression and recognition of informative intent, but not communicative intent? It is plausibly less of a stretch to posit that the expression and recognition of informative intent, alone, always takes place consciously, with effort, and reflectively. Yet even this is problematic. Introspection coheres with the idea that a great deal of communication occurs swiftly, effortlessly, and during other activity placing demands on reflective/executive cognitive processes: in other words, that it is a paradigmatic System 1 ability.

Imagine I am driving home from work and accidently cut off one of my fellow drivers while changing lanes. He leans out the window and utters an imperative sentence which, in line to the post-Gricean picture, causes me to infer his informative intention: that I procreate with myself[87]. However, him inferring the best action to produce so I accurately infer his informative intent is not slow, effortful, and under executive control. Likewise, my inference about his informative intention from his utterance and the context. This is not slow, effortful, and under executive control. The speaker's message seems entirely 'transparent' to me: as transparent as Gibsonians think perception is.

The analogy with perception might be useful here. Just like perception, communication might be 'inferential' without the relevant inferences taking place in central cognition. But if S's inference regarding the best action to produce so I correctly infer his message is not a central process, then in what sense is S expressing an informative *intention*? For intentions are belief-desires pairs, and beliefs and desires are states of central cognition on a Fodorian RTM:

> Belief: '*if I utter X then H will infer my desire that he procreates with himself*'
> Desire: '*I want H to think that he should procreate with himself*'
> Intention: *Utter X!*

---

[87] Or perhaps, via implicature, that I feel insulted.

Likewise, if I am not making inferences about central cognitive states of the speaker, then in what sense am I inferring S's informative *intention*? It seems coherent to posit that the *recognition* of informative intent might occur outside of central cognition. Domain-specific processes in hearers might be equipped to represent domain-general states of speakers. However, it is less coherent to posit that the expression of informative intent (in speakers) takes place outside of central cognition. This is not to say that every kind of internal motivational state is necessarily domain-general and inferentially connected to a suite of other mental states. But if such motivational states are not domain general, then they can't be belief-desire pairs, at least on a Fodorian RTM. In other words they can't be intentional states. This means that speakers aren't always expressing intentions. Concomitantly, it means that hearers aren't always recognising intentions. This is problematic for the post-Griceans and others wishing to explain human communication by appealing to intentions. If a model of communication purports to explain communication by referring to the expression and recognition of intentions, then there is a problem if communication does not always involve the expression and recognition of actual intentions.

The worry is premature, however, as not all central cognition need be System 2-like. Fodor himself (1983, p. 104) thought the bulk of central processing, looking, as it does, "at what the input systems deliver… [looking] at what is in memory, and [using] this information to constrain the computation of 'best hypotheses' about what the world is like", is "largely unconscious". Exposition aside, there is no independent reason to think that *all* central processing must possess System 2 properties, as related by dual-systems theory. Thus, the expression and recognition of intentions relevant to quotidian communication (and communication among young children) might well take place in central cognition, on RTM.

It would be nice to finish off Section 2 here, concluding that post-Gricean and other accounts of communication employing a psychological notion of speaker meaning, are explanatory on RTM. However, we must address a further, important, issue. The issue relates closely to the preceding discussion. Some prominent post-Griceans don't think there *is* a central cognitive system (Sperber & Wilson 2002; Scott-Phillips 2015). The mental architecture envisaged by these theorists is massively modular in nature (e.g. Carruthers 2006). Thus, intention expression and recognition must be the work of domain-specific modules, according to these theorists. But, as we have just seen, the mental states that comprise intentions (namely, beliefs and desires) are not domain-specific mental states, at least paradigmatically. Rather, they are domain-general states possessing inferential relations with each of the subject's other beliefs and desires. This raises at least two questions. First, how can the idea of modular beliefs and desires be made good? Second, how might the expression and recognition of intentions be realised in a wholly modular cognitive system? Answering the first question is relatively simple, resting on the functional role of certain mental representations within a massively-modular architecture. The answer to the second is not as simple. Drawing on Carruthers (2006), I will offer a crude how-possibly model of intention expression and recognition within a massively-modular RTM. The model is intended to show simply that post-Gricean and related forms of speaker meaning are explanatory within a massively-modular RTM. There is no in-principle conflict between being committed to massive modularity, on the one hand, and being a post-Gricean (or one who deflates Gricean speaker-meaning), on the other.

~ Massively modular RTM, intentional psychology, and the expression and recognition of speaker meaning ~

What kind of mental state is a belief within a massively-modular architecture? A belief will be any syntactically-structured mental representation with the right kind of causal/functional role. What is the 'right kind' of causal/functional role, for belief, within a massively-modular architecture? A modular representation will be a belief if it can potentially interact with motivational states in the process of action planning[88]. On Carruthers' (2006, p. 66) preliminary model of a plausible massively-modular architecture of cognition, desires are syntactically-structured representations competing for entry into action-planning modules (of which there are many). When a desire gains entry into an action planning module, the latter queries a restricted range of information-storage modules to devise an action schema to be passed onto the motor control system and drive behaviour. Beliefs are representations in these information-storage modules, used by the action planning system to devise an action schema that might result in the fulfilment of the initial desire. Desires, for their part, are representations triggering action planning systems to query information-storage modules so that the former systems devise action schemas relevant the satisfaction of the initial desires. It is vital to note that, just as there are many action-planning modules, there are many belief-generating and desire-generating modules, with individual action-planning modules connected to one or a limited number of belief-generating modules and desire-generating modules. If there was simply one belief-generating system and one-desire generating system, then the overall architecture would not be very modular-like.

[88] So, what separates beliefs from representations of the world more generally is that the former are available to action planning while phonetic representations (say), or representations of edges in the visual scene (say), are not.

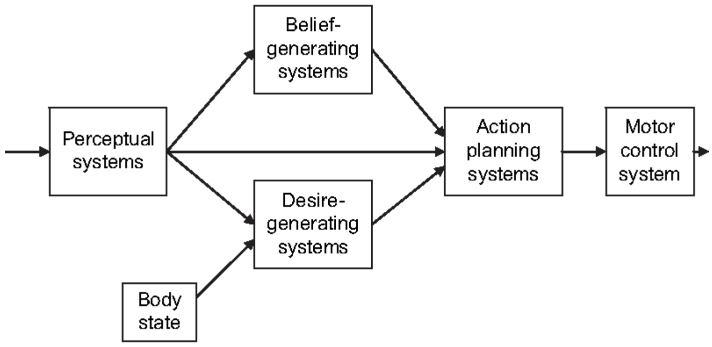

Fig 6.1 A first-pass model of a massively-modular mind without language, from Carruthers (2006).

We can think of intentional states, such as the informative intention that Rick believe the proposition *it's raining*, as a representation in the action planning module that results from an initial desire [for Rick to believe it's raining] triggering the module into operation, causing it to query available belief-generating modules in the process of generating an instrumentally-appropriate action schema (e.g. pointing at the rainy sky for Rick). This action schema is then passed onto the motor control system. Relevant beliefs drawn upon by the initial action-planning module might be memories of how getting people to think that it's raining worked in the past. In addition to representations contained in memory, beliefs can be representations in any conceptual module (such as a mind reading module) that are drawn upon in the production of action schemas by action-planning modules. For instance, the action planning module activated by the desire [that Rick think that it's raining] might query a mindreading module to generate a representation of what Rick will likely infer upon his attention being directed towards the sky.

Of course, when communication involves language, our model will need to be enriched to include modules dedicated to linguistic production and comprehension. Carruthers' (2006, p. 233) more expansive model of the massively-modular mind is presented below in Figure 6.2. Not only are language modules now present, action planning has been expanded to include a practical reasoning module in addition to a module producing action schemas. Moreover, a feedback loop allowing for the mental rehearsal of action schema is included. This will become relevant (later) as a model of how the expression and recognition of intentions might operate reflectively, slowly, and under executive control (i.e. System 2). For now the feedback loop can be ignored.

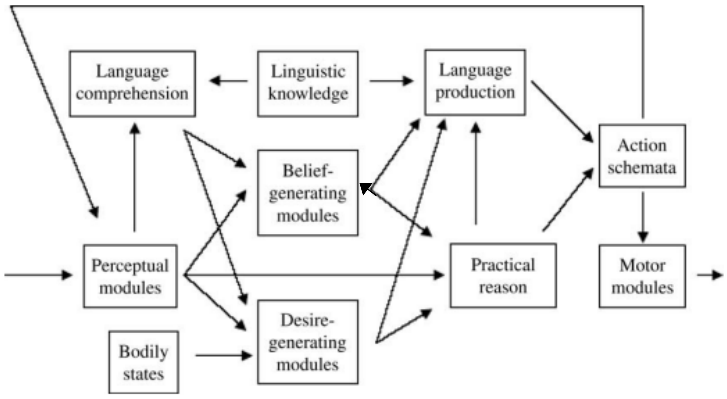

Fig 6.2 A massively-modular model of a mind with language, from Carruthers (2006).

When producing linguistic utterances in the pursuit of one's communicative goals, the core of the process is much the same as outlined previously. One's desire [for Rick to believe that it's raining] triggers a module (this time depicted as a practical reasoning module) querying a restricted range of belief-generating modules to devise a first-pass action schema so that Rick will believe it is raining. What is new, this time around, is that linguistic knowledge is used, in addition to memories and mindreading information (i.e. information in belief-generating modules). The flow of information beginning with the practical reasoning module and ending in action schema generation goes through the language production module. As a result, the action schema produced is linguistic in nature (the utterance 'Look! It's raining!').

This is how informative intentions might be produced, causing communicative behaviours in speakers in a way which is fast automatic and relatively effortless (i.e. System 1). But how might the recognition of informative intentions go, in hearers? And, additionally, how might informative intentions be expressed and recognised in System 2 fashion? For, despite most communication being relatively unreflective and automatic, there *are* instances where speakers devote great care and cognitive effort to expressing their meaning, and where hearers devote great care and effort in inferring speaker meaning. First, the recognition of informative intent in hearers.

The process starts with the hearer's perceptual modules encountering visual or verbal stimuli of the speaker's pointing gesture and/or utterance. Take the non-verbal signal first. As with all perceptual information, visual information about the speaker's pointing gesture is globally broadcast to the whole suite of conceptual (belief and desire) modules. If this perceptual information falls within the input domain of a conceptual module, the latter will be activated. One module(s) kicked into gear will be the hearer's mindreading modules(s). Mind reading modules will be activated upon the perception of any kind action carried out by another agent, whether communicative or not, in order to determine the agent's goal. However, if the speaker has produced signals indicative of communication (Csibra 2010), then one of the conceptual modules kicked into gear might be the hearer's relevance submodule (Sperber & Wilson 20002), which, if the speaker acted in line with the principle of relevance, enables the hearer to infer the speaker's informative intention/meaning. The hearer's representation of the speaker's informative intention is then stored as a belief (in the massively-modular sense) in memory. When the speaker's communicative action is linguistic, the process will, in addition, involve language modules. Perceptual information about the speaker's verbal behaviour is passed to the language comprehension module, which then queries linguistic knowledge, in addition to the mindreading module(s) and relevance sub-module, before issuing a belief representing the speaker's meaning (or something close enough).

As noted just above, some episodes of communication are not System 1 processes. Sometimes, speakers devote great care and cognitive effort in expressing their meaning, such as when my partner writes a breakup letter she leaves on the kitchen bench. Likewise, hearers sometimes devote great care and effort in inferring speaker meaning, like when I pour over each sentence in the breakup letter. How might a massively-modular architecture enable reflective, System 2, forms of communication? This is where the feedback loop in our model of the massively-modular mind enters in picture. According to Carruthers (2006), System 2 reasoning is not the product of a single, domain-general, cognitive mechanism, like Fodor's central system. Instead, System 2 processing is realised by cycles of domain-specific modular cognition: the creation and mental rehearsal of action schema, where many of these actions are linguistic: i.e. episodes of 'inner speech'[89].

Action schema rehearsal takes place when practical reasoning issues intentions to one of the individual's action schema modules, but where this action schema is then fed back to perceptual modules instead of forward to motor control modules. Since perceptual modules broadcast their information to the whole suite of conceptual modules, rehearsed action schemas can trigger conceptual modules not responsible for initially creating the action schema. Sometimes, the language production system is queried in the creation of offline action schemas. This results in episodes of inner speech. Inner speech is a particularly powerful form of action schema rehearsal, on Carruthers' model. This is because language production can conjoin representations generated in a range of different conceptual modules. What results is a verbal action schema that cuts across domains, and which is then fed back into the entire cognitive system to be processed by the various domain-specific modules.

[89] Although not exclusively: the mental rehearsal of visual imagery plausibly plays a role in System 2 thinking.

How might the System 2 expression of informative intentions work? Take the utterance “your eyes are like sapphire”, spoken to your date. Imagine your utterance is the result of explicit reflection aimed at influencing your date to kiss you. That is, your informative intention is not to instil the belief in your date that her eyes are literally like sapphire. Instead, your informative intention is the initiation of a kiss. We can model the process as follows. First, an initial verbal action schema is created off the back of a desire to kiss. This involves the desire to kiss triggering your practical reasoning module to consult a limited range of conceptual, belief-generating, modules as well as your language production module. A verbal action schema might then be chosen, for instance, “can I kiss you?”. However, instead of being passed to motor modules, and being uttered out loud, the (verbal) action schema of this utterance is taken offline and fed back to your perceptual modules. As a result, it is globally broadcast to all your conceptual modules, a process that is experienced as an episode of inner speech.

The offline rehearsal of verbal action schemas, or inner speech, provides a plausible mechanism for how you could predict the likely consequences of saying “can I kiss you?”. Conceptual modules, such as mindreading, receive a representation of this utterance, and generate, for instance, the prediction that you would find this too direct. This then sets off a whole new cycle of action planning. This process might be repeated *n* number of times before the verbal action schema “your eyes are like sapphire” is sent to your motor modules to be uttered out loud. What the offline rehearsal of verbal action schemas contributes to cognitive processing is not only multiple cycles of modular processing, and thus planning refinement, but also the recruitment of conceptual modules not involved in the initial cycle of action planning (the one responsible for the verbal action scheme asking directly for a kiss).

What about the System 2 recognition of informative intentions? On Carruthers’ model, your date perceives your utterance and this information is fed to her language comprehension module. After a semantic representation of the literal meaning of your utterance is generated there, it is fed to the conceptual modules, including mindreading and the relevance sub-module within mindreading. Now, assume that, for some reason, mindreading and relevance cannot deliver a clear verdict. System 1 inference of your meaning is not able to infer a clear informative intention with just one cycle of modular processing. Is it your intention just to complement, or something more? As a result, a verbal action schema is generated: “What on earth does ‘your eyes are like sapphire’ mean?!”. This is fed back to the perceptual modules and globally broadcast to the whole suite of conceptual modules after being decoded by the linguistic comprehension module. This time, long term memory is queried for information. Perhaps a memory of a movie depicting a very similar situation is recalled. As a result, a belief is generated in your date that, in fact, you desire a kiss.

It has been my goal to provide but the barest sketch of how the expression and recognition of informative intentions might explain communicative behaviour, according to RTM. I have used Carruthers' model of the massively-modular mind, because it has been worked out in some detail, and because prominent post-Griceans seeking to explain how speakers and hearers express and recognise meaning via intentional psychology favour a massively-modular mental architecture. There might be many problems with massive-modularity, as a hypothesis about how the mind works; or with Carruthers' particular model; or with my application of it. However, the point has been simply to show that the expression and recognition of intentions is, at least in-principle, explanatory of some crucial aspects of communication: namely, the expression and recognition of speaker meaning. Having done this, I now consider the explanatory status of speaker meaning assuming the relationship between intentional psychology and cognitive science is more nuanced than depicted on RTM.

3. Assuming the intentional stance

I argued that the substance of post-Gricean accounts of communication depend on the explanatory status of intentional states. I explored some different accounts of the explanatory status of intentional states: i.e. ways in which folk psychological explanation, invoking intentional states, interfaces with explanation in the cognitive sciences. One account drew from the RTM and took intentionality to be grounded in syntactically-structured beliefs and desires, these being type-type reducible to cognitive states. On this view, post-Gricean accounts of communication are explanatory. They are explanatory because the kinds of folk states central to post-Gricean accounts of communication neatly 'hook onto' kinds of states that are explanatory from the perspective of cognitive science.

The other relevant account of the ontology of intentionality to be explored is the intentional stance (Dennett 1987). It takes intentionality to be grounded in the right kinds of behavioural patterns, as opposed to internal mental states (Dennett 1992). Note that the understanding of 'explanation' employed above was decompositional: appealing to how sub-functions give rise to system-level functions. From this, it seems natural to conclude that a Dennettian view of the interface problem is not an option for post-Griceans. This is because, on the intentional stance, the beliefs and desires constituting intentionality are not discrete states of the cognitive system. Why is this relevant? Well, according to one attractive view, there are two requirements of person-level psychological explanation (PPE) (Bermudez 2005). First, PPE depends on the existence of regularities governing the inter-relations between person-level states and between these states, perceptual representations, and behaviour. Second, PPE depends on causally efficacious internal items corresponding to the person-level states cited in PPEs. From this second requirement, one might infer that the intentional stance is not an option for post-Griceans who wish to cognitively explain how meaning is expressed and recognised in human communication.

Consider also one of Dennett's motivations for advocating the intentional stance. The goal is to insulate person-level explanation from developments in cognitive science. Whether the subpersonal basis for cognition turns out to consist in representations bearing a structural correspondence with linguistic representations of the propositional attitudes, or whether what's under the hood instead consists in distributed neural networks (or whatever), bearing no neat correspondence with person-level states, PPEs will nevertheless be truth-apt[90] explanations of behaviour. They will be truth-apt explanations of behaviour because they potentially correspond, not to internal, causally-interacting physical structures, but rather to genuinely existing patterns in the outward behaviour of agents. The truth of the claim 'x believes that *p*' depends, not on states internal to x, but rather to x's behaviour – either actual or counterfactual. The structural correspondence between PPE and the world goes from being an internal correspondence (on RTM) to an external one. Thus, one could easily conclude that the intentional stance is not explanatory in a decompositional sense.

If the intentional stance is not a way of decomposing an intelligent system into its sub-systems, then what is it? How does it explain? It explains by positing a network of relations between inputs, mental states, and behaviours; where intervening on any one of these variables makes a difference to the others. According to Dennett (1992, p. 43), "if one finds a predictive pattern of the sort just described one has ipso facto discovered a causal power – a difference in the world that makes a subsequent difference testable by the standard empirical methods of variable manipulation". The idea is that, taking the intentional stance, PPE is explanatory because whether 'x believes that *p'* is true depends on what x *would* do, say, or think in various circumstances: "what it means to say that someone believes that *p*, is that that person is disposed to behave in certain ways under certain conditions" (Dennett 1987, p. 50). And these counterfactuals can be tested, either via observation in the wild or experimentally.

Take the two sentences below. They both explain, from the perspective of the intentional stance. Each are made true or false, depending on counterfactual behaviours, utterances, and thoughts.

---

[90] I.e. not just useful. Dennett arguably became less of an instrumentalist by the time of *Real Patterns* (Dennett 1992).

1. *S intends that H recognise S intends H to go away* (i.e. S harbours communicative intent)
2. *S intends that H go away* (i.e. S harbours informative intent)

Consider the first sentence. In the last chapter I discussed a study probing the truth of relevant counterfactuals to determine whether sentence 1 is true or false (Grosse et al. 2010). To briefly recap, the study required infants to request a desired object from a first experimenter (say, a ball to complete a task the infant was playing with a second experimenter). The first experimenter would then 'misunderstand' the infant's request for the ball by pointing to a different object (e.g. a bit of paper), uttering 'Oh, you want the paper!'. However, at the same time as reaching for the undesired object, the first experimenter would distractedly place the requested object (the ball) in easy reach of the infant, so that the infant's informative intention was essentially satisfied despite being misunderstood. This 'misunderstanding' condition was compared with a series of other conditions: most notably where the infant got the ball *and* was understood correctly. It was found that around 45 per cent of 18 month-old infants attempted to 'repair' the communicative interaction in the misunderstanding condition. It was also found that, when an infant got what it wanted but was misunderstood, the time taken to turn away from the communicative exchange was a bit more than double the length of time (around 9 seconds) compared to when the infant got what it wanted and was understood correctly (around 4 seconds). The conclusion of the study is that infants communicate with communicative intent over and above informative intent.

Considering this, the difference between RTM and the intentional stance isn't that the former is explanatory while the latter is not. Rather, the difference comes down to the kind of 'explanation' given by invoking person-level states. On RTM, (at least many) person-level states are type-type reducible to cognitive states. For instance, my intention to lean back for Rick is type-type reducible to a state in one of my practical reasoning modules, after this module has consulted belief-generating modules in response to being activated by my desire that Rick believe Roger is approaching. Looking back at Figures 6.1 and 6.2, it is easy to come away with the impression that a mechanism is being described, where intentional states are components in this mechanism. Indeed, proponents of mechanistic explanation argue that cognitive models like these explain only to the degree they depict mechanistic structure (Kaplan & Craver 2011, Piccinini & Craver & 2011). Others have objected, arguing that many cognitive models represent, not mechanistic structure, but rather functional organisation (Weiskopf 2011, Barrett 2014, Shapiro 2016). Unless mechanistic explanation is trivialised into a vague, catch-all, physicalism, a cognitive model is not mechanistic. While all cognitive scientists expect their models to ultimately bottom-out in physical structures, many test their hypotheses about cognitive organisation without considering the brain at all. What these cognitive scientists include in their models are components described functionally, as opposed to mechanistically. Functional decomposition investigates how sub-functions (i.e. a practical reasoning module) contribute to some system-level function (i.e. expressing meaning); mechanistic decomposition investigates how physical structure (shape, position, timing, force, charge) produces the activities the functionalist identifies (Shapiro 2016). Taking this line, person-level states, on RTM, are states of a system that is functionally decomposed. Functional decomposition is similar to mechanistic decomposition in that both decompose a system into smaller units which, when combined, give rise to a higher-level function. However, functional decomposition is concerned less (if at all) with how physical structure realises function.

The idea I want to take forward is not the difference between mechanistic explanation and functional decomposition, per say. Instead, it is the idea that much cognitive science explains behaviour via decomposition. This contrasts with the intentional stance. The intentional stance explains behaviour by highlighting counterfactual dependencies at the level of the whole system. Having pointed out that the intentional stance doesn't decompose intelligent systems to understand them, though, what are we to make of Dennett's (1987) suggestion that his intentional stance is like Marr's (1982) top level in the latter's famous hierarchy of cognitive theorising? Marr calls his top level the *computational* level, distinguishing it from both the (middle) *algorithmic* level and the (bottom) *implementation* level. A computational description of some thinking/behaving system is the most abstract. It involves characterising the system's inputs and outputs, and the relation between these inputs and outputs. The latter relation constitutes the *function* (in the mathematical sense of the term) the system can be said to compute, which it is then the job of the algorithmic level of analysis to specify computational processes that can realise the function defined at the computational level. The implementation level involves physical realisation of the processes, formulated at the algorithmic level, in neural hardware (or silicon chips). For Marr at least, a good computational (i.e. top-level) analysis of some system is key to further investigation down the hierarchy. If, as Marr took to be the case in then-contemporary vision science, the computational description of some capacity is mistaken, then work at lower levels will be misguided.

Now, if taking the intentional stance towards a system is akin to analysing the system at Marr's computational level, then perhaps the former will count as the *beginnings* of a decompositional explanation of behaviour. The job of the intentional stance will be to get clear on the capacities of some system, before programmers or network engineers get to work designing computational implementations of these capacities. This sounds tantalising. However, despite comparing the intentional stance with Marr's computational level, Dennett himself holds-out little hope in finding a "strong resemblance" between an intentional depiction of some system (what he also calls a 'competence model') and a depiction of the same system at levels corresponding to Marr's algorithmic and implementation analyses.

> The fact about competence models that provokes my "instrumentalism" is that the decomposition of one's competence model into parts, phases, states, steps, or whatever *need* shed no light at all on the decomposition of the actual mechanical parts, phases, states or steps of the system being modelled-even when the model is, as a competence model, excellent (Dennett 1987, pp. 75-76).

If indeed there is a resemblance between an intentional description of some system, on the one hand, and a description of the system from what Dennett calls "sub-personal cognitive psychology", this resemblance will be more or less accidental. Moreover, it won't be indicative of a 'good' intentional description. What is indicative of a good intentional description is, instead, whether the attributed person-level states: a) make sense of the system's behaviour, or b) accurately predict, again constrained by considerations of rationality, what the system will do.

There are at least two conclusions stemming from the forgoing considerations. The first privileges explanation via decomposition and, as a result, disprivileges the explanatory status of the intentional stance. Because the latter doesn't decompose an intelligent system into its lower-level sub-functions, it doesn't give us much understanding of the system's intelligence. I favour a second conclusion. Intentional psychology (PPE from the perspective of the intentional stance) is obviously a powerful way of intervening on an intelligent system for the purposes of manipulation and control. As a result, it would be an exercise in extreme parochialism to dismiss intentional psychology simply because it doesn't capture regularities at the level of internal cognitive structure. Forms of empirical clinical psychology, like cognitive behavioural therapy and acceptance and avoidance therapy, are saturated in person-level states: for example, identifying and then working through one's (possibly irrational) beliefs surrounding perceived negative problems. From the perspective of the intentional stance, the relationship between inputs, mental states, and behaviours is a systematic one, such that intervening on, say, one's un-analysed and perhaps irrational belief about the expectations of partners within marital relationships, has (possibly positive) consequences for other mental states (including emotions) and for behaviour. Consider also any empirical study in psychology/cognitive science based on the violations-of-expectations paradigm. For instance, in response to a child's surprise at where Sally looks for her marble after returning from her 'walk', it is concluded the child lacks certain beliefs: in this case, veridical beliefs about Sally's (false) mental states. This conclusion is made assuming a systematic relationship between behaviours (e.g. surprise and gaze direction) and mental states (beliefs about Sally's mental states). The conclusion that the child lacks certain beliefs (or the capacity to form certain beliefs) is also projectable to future behaviours. This is, in effect, what is being done when the child is diagnosed with some form of mind-blindness as a result of failing the false-belief test.

Now, one objection to this is Fodorian. Something like the RTM *must* be true, not only because of the success of PPE in allowing us to navigate our peers every day, but also because successful science in psychology takes place on the level of PPE. This isn't the only available conclusion, though, and begs the question against the intentional stance. The intentional stance claims a system of regularities holding at the level of PPE, without being committed to regularities internal to intentional systems resembling those holding at the level of PPE. My point is that, if there is indeed a lack of resemblance between PPE and internal cognitive structure, then it is foolish to demote PPE because it doesn't capture internal structure in the process of successfully exploiting structure at the level of PPE. Psychological explanation is many and varied. Sometimes, i.e. when working out how to build an artificial system that mirrors human intelligence, it may be necessary to engage in functional and mechanistic decomposition. At other times it won't be, when manipulating thought and behaviour at an abstract level.

When, then, of models of communication invoking intentional states? If the intentional stance is the proper response to the interface problem, then such models won't be explanatory in a decompositional sense. However, they *will* be explanatory in an interventionist sense. In leaning back on my chair, the claim that I behave motivated by the informative intention that Rick believe Roger approaches will be literally true or false. Whether it is true or false depends, for instance, on what I would do/think/say in response to Rick exclaiming 'Yay!', after Rick has looked over his right shoulder. Assume that I believe Rick loathes Roger, and that Rick desires not to interact with Roger. Then, if I genuinely intend Rick to believe Roger is approaching, I should be surprised at a gleeful response from Rick. Additionally, if the intentional stance is the proper response to the interface problem, then positing that speaker meaning is expressed and recognised *via* the expression and recognition of certain intentions might be a true explanation of communication. It might literally be the case that, in order to work out what you mean, I infer your intention(s) in gesticulating towards me. If PPE is the kind of intentional 'calculus' Dennett suggests, then even if personal-level states don't neatly correspond to states of subpersonal cognition, it nevertheless obtains that, by (correctly) inferring that you want me to go away, I will have a good chance at successfully coordinating my behaviour with yours: at least, a much better chance than if I (incorrectly) infer you want me to come closer.

4. Conclusion

Post-Gricean accounts of communication are hostage to how descriptions of the mind at the person-level (invoking intentional states) *interface* with descriptions of mind at lower levels. I presented some ways of conceiving of this relationship among levels of description, and explored the ramifications for post-Gricean accounts. I concluded as follows. First, if RTM accurately depicts the relationship between person-level states and cognitive states, then post-Gricean accounts of communication come out as explanatory in a decompositional sense. Second, if this relationship is as Dennett theorised, then post-Gricean models still explain how expressing and recognising meaning in communication works: this time at the level of whole persons.

Conclusion

Against the background of the information vs influence debate (Owren, Rendall & Ryan 2010, Stegmann 2013), this thesis defended a broadly informational approach to animal communication. However, it sought to get clear on exactly what cognitive commitments arise from attributing meaning to signals, as well as whether animal communication can be defined in terms of such meaning. I drew a line between a working notion of 'natural meaning', requiring very little cognitive sophistication; and a notion of 'speaker meaning', requiring intentional behaviour on the part of speakers and mindreading on the part of hearers.

Chapter 1 argued that the natural meaning of signals should be grounded by both correlation and by teleology. Only if so grounded can the content of signals do the following: (1) make sense of why receivers currently respond to signals, (2) explain historical success non-vacuously, and (3) assign the appropriate contents in non-cooperative cases. Chapter 2 argued that, despite the explanatory upshot of attributing natural meaning to signals, signalhood cannot be defined in terms of phenotypes that uniquely carry natural meaning. This conclusion was not an artefact of the particular notion of natural meaning I defended in Chapter 1. It would be the result, also, of a purely correlational notion or a purely teleological one. Chapter 3 then looked more closely at the role of natural meaning when it comes to explaining receiver responses. What do (or should) ethologists mean by saying that a creature behaved thus-and-so because it derived natural meaning *q* from a signal? Typically, this explains a receiver's (relatively fixed) response from a historical and population-level perspective only. Sometimes, though, it is part of explaining a receiver's (relatively flexible) response in real time.

Chapter 4 switched focus from natural meaning to speaker meaning. This was done because non-natural meaning is sometimes attributed to the signals of a limited number of nonhuman species, such as chimpanzees. I tried to make clear sense of non-natural meaning while outlining its psychological commitments. Chapter 5 argued that these commitments might well place non-natural meaning outside of the capabilities of nonhuman species, if not also young human beings learning how to communicate for the first time. I suggested that the latter point rests on whether or not children possess System 1 recursive mindreading skills from birth: a question in need of further empirical investigation. Drawing from Csibra (2010), I also suggested there are grades of speaker meaning standing in-between natural meaning, on the one hand, and non-natural meaning, on the other. Both intermediary forms of speaker meaning involve receivers inferring informative intent, but not communicative intent. On the lower end, there might be no proxy yet of communicative intent (i.e. chimpanzee communication). On the upper end, however, there might be a proxy for the expression and/or recognition of communicative intent: codes that trigger interpretive processes in hearers in virtue of proper function as opposed to meta-representation.

Finally, Chapter 6 explored the relationship between intentional psychology and cognitive science. I did this because notions of speaker meaning (whether fully Gricean or not) are grounded by intentional states. I considered the ramifications of two different ways intentional psychology might interface with cognitive science. I concluded that the ramifications are different depending on one's solution to the interface problem. On neither solution to the interface problem, however, should speaker meaning be rejected as non-explanatory.

Before finishing, I offer some brief concluding remarks about the project and its underlying assumptions. These would have taken up too much space in the introduction. Furthermore, they might not have made sense there, without the context of the thesis behind them.

- More on terms -

Those who are familiar with the philosophy of information/representation may take issue with the way I used the notion of 'natural meaning'. Some philosophers, particularly Millikan (2004) and Neander (2017), use this term differently. According to them, 'natural meaning' is incapable of being false. This is because it is carried solely in virtue of a sign being physically linked to its referent. *M* number of rings on a tree naturally means the tree is *n* years old, because of the physical link between tree rings and tree age. Unless the tree is *n* number of years old, you will not see *m* number of tree rings. This means signs carrying natural meaning can't misrepresent. Philosophers taking this line have contrasted natural meaning with a kind of meaning that, they allege, *can* misrepresent.

Somewhat confusing, in my opinion, Piccinini & Scarantino (2010) contrast the kind of (natural) meaning carried by tree rings, which they also allege can't be false, with Gricean non-natural meaning, which allegedly can. They then state the latter kind of meaning is "often used interchangeably with the notion of representation" (Piccinini & Scarantino 2010, p. 242). Now, it is true that Millikan and Neander (and others) contrast natural meaning with representational content, because the latter can be false while, allegedly, the former cannot. However, the latter philosophers do not equate non-natural meaning with representational content. And they are right not to. Non-natural meaning is sent by individuals behaving with higher-order intentions towards hearers, and it is received by individuals able to meta-represent the higher-order intentions of senders. Representational content, on the other hand, does not require anything like this. Take a subpersonal mental representation in the early visual system. This carries representational content without being sent by something that is even capable of behaving motivated by intentions; and this content can be 'received' without the consumer having the capacity to recognise intentions.

Although I stand with Millikan and Neander on keeping Gricean non-natural meaning separate from representational content, I part ways with them in attributing, as I do, natural meaning to signals. As I have been using the notion of 'natural meaning', signals carrying it *can* misrepresent. Take our female *Photuris* firefly offering sex but instead delivering death. I have understood her signals to be misrepresenting that she is a *Photinus* female looking for a mate. But for Millikan and Neander, natural meaning can't misrepresent. This is because, for them, natural meaning (as opposed to representational content) rests on causation/correlation alone. The transition from natural meaning to representational content, for them, requires the introduction of biological proper function. Only after we add proper function can a sign that correlates with its referent possibly be false (and only then can it become a genuine *sign*). On the other hand, I have used the term 'natural meaning' to refer to signals that *can* misrepresent, because of biological proper function.

I did this, firstly, because I wanted a neat way of contrasting Gricean non-natural meaning with a simpler kind(s) of meaning not requiring the expressing and recognition of intentions. Contrasting non-natural meaning with, say, 'content' or 'information', as opposed to 'natural meaning', is not as symmetrical. Secondly, I used the term 'natural meaning' to refer to a kind of meaning that can misrepresent because, at least from my perspective, the latter is much closer to 'orthodox' natural meaning - as Millikan and Neander use the term - than non-natural meaning. Combining correlation with proper function, in order to turn a concept of natural meaning that (allegedly) cannot misrepresent into one than can misrepresent, is a relatively minor adjustment to the concept of 'natural meaning'. A third reason for my use of 'natural meaning' is that 'orthodox' Gricean natural meaning *can* misrepresent, if we allow the latter to be carried by entities that correlate with their referents, as opposed to demanding causation (e.g. Skyrms 2010). This is because the correlation grounding content exists at the level of types. As a result, tokens of a type R that correlates with C can be tokened when C does not obtain (Shea 2007). Thus, the reason for splitting natural meaning from representational content may not be a good reason to start with.

But does my use of terminology have the following, awkward, consequence? Am I forced to say that the information in cues, for instance, isn't natural information? One might think I am, because my notion of natural meaning requires biological proper function (on the part of consumers of this meaning). The particular natural meaning carried by a signal refers to the world-state that rationalises the evolved/learnt response of the receiver, in addition to referring to the world-state which the signal correlates with. However, because the theory of content chosen in Chapter 1 doesn't require co-adapted producers and consumers, but only adapted consumers, cues also carry natural meaning, as I have been using that term. If a cue is used by an organism to adjust behaviour to the world, the organism's response must be a product of evolution or learning. Thus, there will be a world-state that rationalised the response of an organism to a cue when, historically, things went well for that organism and the response was re-enforced. This world-state is the natural meaning carried by the cue. Of course, cues that have not yet been utilised by an organism won't carry natural meaning. However, they will still correlate with their potential referents. A beam of light reflected by a predator correlates with the predator, even if this cue has not yet been utilised, and this correlation explains why receivers that eventually came to respond appropriately to the cue benefitted.

- *Is all meaning natural meaning?* –

Finally, I respond to Skyrms' (2010, p. 1) provocative claim that all meaning is natural meaning. In his book, Skyrms presents a how-possibly account of the emergence of meaning conventions from the spontaneous actions of individuals in a population. Lewis (1969) first proposed something along these lines. He did so in response to the following conundrum. How do words get their semantic contents? It can't be from some episode where everyone sits down to decide upon the semantic contents of words, because this episode would itself depend on everyone possessing contentful language. Instead, Lewis proposed that semantic content arises through the spontaneous coordination of senders and receivers. However, the coordination of senders and receivers envisaged by Lewis depends on them possessing a raft of (contentful/meaningful) mental states. Only by me *believing* that others in my language community use words in a certain way (and that, additionally, these other people believe that I believe this) do I respect semantic conventions.

Skryms (2010) took the idea of spontaneous coordination giving rise to meaning conventions a step further than Lewis, proposing that this coordination can be achieved on basis of trial and error (either evolutionary or ontogenetic) as well as via the rational coordination of agents. With the right kind of feedback, the spontaneous behaviours of 'mindless' individuals (like fireflies) can be coordinated such that meaning conventions arise: senders will produce behaviours in certain world-states, such that this behaviour influences receivers, causing the latter to produce their own behaviours that are adaptive (enough). The resulting 'meaning' that emerges from this 'mindless' coordination is correlational. Signals carry natural meaning about particular world-states in virtue of raising the probability of such world-states. Thus, we have meaning conventions grounded by facts, none of which are facts about meaningful entities prior to the emergence of meaning conventions. Meaning conventions can arise on the back of non-mentalistic coordination. As a result, "all meaning is natural meaning" (Skryms 2010, p. 1).

This might be well and good for communication in many, relatively simple, species. But what about primate (including human) communication? Chimpanzee communication arguably involves senders behaving with informative intentions and receivers recognising these intentions. Quotidian human communication involves the same, perhaps with the aid of assumptions constraining the process of inferring intent that are triggered by ostensive signals (Csibra 2010). If this is true, then the idea that all meaning is natural meaning is not as straight forward as Skyrms makes out. For, the kind of (speaker) meaning humans, and potentially other great apes, exchange depends on the intentional states of interlocutors; where these states possess intentionality, or 'aboutness', independent of external semantic conventions.

If Skyrms's claim is to be born out, it would have to be shown that the intentional states grounding speaker meaning themselves carry merely natural meaning. This brings us back to the interface problem, discussed in Chapter 6. If, for example, the intentional states grounding the expression and recognition of speaker meaning are discrete mental representations in the vein of a representational theory of mind, then Skyrms's claim that all meaning is natural meaning depends on the success of the long-running project to 'naturalise' the content of internal mental representations[91]. If discrete mental representation indeed underpin propositional attitude ascriptions, and if the intentional content of these representations (at least at the base level) can be grounded by appeal to the information carried by such representations (Dretske 1981) or by teleological theories (Millikan 2004), then the following picture of communication is suggested: linguistic utterances carry natural meaning about the intentions of speakers, where these intentions themselves carry natural meaning qua mental representations, in the sense outlined by Dretske, Millikan or someone similar. Alternatively, if the interface problem is solved by taking the intentional stance, then linguistic utterances carry natural meaning about the

[91] See Stich & Warfield (1994) for an overview of this literature.

intentions of speakers, where these intentions are simply mental/behavioural dispositions of speakers, as opposed to internal mental representations corresponding to propositional attitude ascriptions. In either case, all meaning (external *and* internal) would indeed be natural meaning, but in a way that is somewhat more nuanced than the Skyrmsian proclamation. Of course, such a picture would need to be filled out in much more detail than I have done. As such, I end merely with the aforementioned suggestion for how natural meaning and speaker meaning might combine to form a unified account of communication in which a) natural meaning is primary, but where b) speaker meaning is taken seriously.

References

Allen, C. & Hauser, M.D. 1991. 'Concept attribution in nonhuman animals: theoretical and methodological problems in ascribing complex mental processes', *Philosophy of Science*, vol. 58, no. 2: 221-240.

Allen, C. & Bekoff, M. 1997. *Species of mind: the philosophy and biology of cognitive ethology*, MIT Press: Cambridge, MA.

Apperly, I.A. & Butterfill, S.A. 2009. 'Do humans have two systems to track beliefs and belief-like states?', *Psychological Review*, vol. 116, no. 4: 953-970.

Ariew, A. 2003. 'Ernst Mayr's 'ultimate/proximate' distinction reconsidered and reconstructed', *Biology & Philosophy*, vol. 18, no. 4: 553-565.

Arruda, C.T. & Povinelli, D.J. 2016. 'Chimps as secret agents', *Synthese*, vol. 193, no. 7: 2129-2158.

Artiga, M. 2014. 'Signaling without cooperation', *Biology & Philosophy*, vol. 29, no. 3: 357-378.

Barrett, D. 2014. 'Functional analysis and mechanistic explanation', *Synthese*, vol. 191, no. 12: 2695-2714.

Bermudez, J.L. 2003. *Thinking without words*, Oxford University Press: Oxford, UK.

Bermudez, J.L. 2005. *Philosophy of psychology: a contemporary introduction*, Routledge: New York.

Birch, J. 2014. 'Propositional content in signalling systems', *Philosophical Studies*, vol. 171, no. 3: 493-512.

Blackburn, P.L. 2007. *The code model of communication: a powerful metaphor in linguistic metatheory*, SIL International: Dallas.

Bloom, P. 2000. *How children learn the meanings of words*, MIT Press: Cambridge, MA.

Botterill, G. & Carruthers, P. 1999. *The philosophy of psychology*, Cambridge University Press: Cambridge.

Braddon-Mitchell, D. & Jackson, F. 1996. *Philosophy of mind and cognition*, Blackwell: Oxford.

Brownstein, M. 2014. 'Rationalizing flow: agency in skilled unreflective action', *Philosophical Studies*, vol. 168, no. 2: 545-568.

Bruner, J.P., Brusse, C., Kalkman, D. 'Cost, expenditure and vulnerability', *Biology & Philosophy*, vol. 32, no. 3: 357-375.

Burge, T. 2010. *Origins of objectivity*, Oxford University Press: Oxford, UK.

Butterfill, S.A. & Apperly, I.A. 2013. 'How to construct a minimal theory of mind', *Mind & Language*, vol. 28, no. 5: 606-637.

Camp, E. 2007. 'Thinking with maps', *Philosophical Perspectives*, vol. 21, no. 1: 145-182.

Carazo, P. & Font, E. 2010. 'Putting information back into biological communication', *Journal of Evolutionary Biology*, vol. 23, no. 4: 661-669.

Carruthers, P. 2006. *The architecture of the mind*, Oxford University Press: Oxford, UK.

Cheney, D.L. & Seyfarth, R.M. 1988. 'Assessment of meaning and the detection of unreliable signals by vervet monkeys', *Animal Behaviour*, vol. 36, no. 2: 477-486.

Cheney, D.L. & Seyfarth, R.M. 2007. *Baboon metaphysics*, University of Chicago press: Chicago.

Csibra, G. 2010. 'Recognising communication intentions in infancy', *Mind & Language*, vol. 25, no, 2: 141-168.

Davies, N.B. & Halliday, T.R. 1978. 'Deep croaks and fighting assessment in toads Bufo bufo', *Nature*, vol. 274: 683-685.

Dawkins, R. & Krebs, J.R. 1978. 'Animal signals: information or manipulation', in J.R. Krebs & N.B. Davies (eds), *Behavioural Ecology: an evolutionary approach*, Blackwell: Oxford, pp. 282-309.

Deacon, T. 1997. *The symbolic species*, Norton: New York.

Dennett, D.C. 1987. *The intentional stance*, MIT Press: London.

Dennett, D.C. 1991. *Consciousness explained*, Little Brown & Co: New York.

Dennett, D.C. 1992. 'Real patterns', *Journal of Philosophy*, vol. 88, no. 1: 27-51.

Dennett, D.C. 1998. *Brainchildren*, MIT Press: Cambridge, MA.

Dickins, T.E. & Barton, R.A. 2013. 'Reciprocal causation and the proximate-ultimate distinction', *Biology & Philosophy*, vol. 28, no. 5: 747-756.

Dretske, F.I. 1981. *Knowledge and the flow of information*, MIT Press: Cambridge, MA.

Drestke, F.I. 1986. 'Misrepresentation', in R. Bogdan (ed), *Belief*, Oxford University Press: New York: 17-36.

Dretske, F.I. 1994. 'The explanatory role of information', *Phil. Trans. R. Soc. Lond. A,* vol. 349, no. 1689: 59-70.

Evans, J.S.B. 2008. 'Dual-processing accounts of reasoning, judgment, and social cognition', *Annu. Rev. Psychol.*, vol. 59: 255-278.

Fitch, W.T. 2010. *The evolution of language*, Cambridge University Press: New York.

Fitch, W.T. & Zuberbuhler, K. 2013. 'Primate precursors to human language: beyond discontinuity', in E. Altenmuller, S. Schmidt & E. Zimmerman (eds), *Evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man*, Oxford University Press: UK: 26-48.

Fodor, J.A. 1975. *The language of thought*, Harvard University Press: Cambridge, MA.

Fodor, J. A. 1983. *The modularity of mind*, MIT Press: Cambridge, MA.

Fodor, J.A. 1996. 'Deconstructing Dennett's Darwin', *Mind & Language*, vol. 11, no. 3: 246-262.

Gallistel, R.C. 2003. 'Conditioning from an informative processing perspective', *Behavioural Processes*, vol. 62, nos. 1-3: 89-101.

Gergely, G. & Csibra, G. 2003. 'Teleological reasoning in infancy: the naïve theory of rational action', *Trends in Cognitive Sciences*, vol. 7, no. 7: 287-292.

Godfrey-Smith, P. 1992. 'Indication and adaptation', *Synthese*, vol. 92, no. 2: 283-312.

Godfrey-Smith, P. 1996. *Complexity and the function of mind in nature*, Cambridge University Press: Cambridge.

Godfrey-Smith, P. 2000. 'On the theoretical role of "genetic coding"', *Philosophy of Science*, vol. 67, no. 1: 26-44.

Godfrey-Smith, P. 2009. *Darwinian populations and natural selection*, Oxford University Press: UK.

Godfrey-Smith, P. 2013. 'Information and influence in sender-receiver models, with applications to animal behaviour', in U.E. Stegmann (ed), *Animal communication theory: information and influence*, Cambridge University Press: Cambridge: 377-393.

Godfrey-Smith, P. & Sterelny, K. 2016. 'Biological Information', *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), E.N. Zalta (ed), URL = <https://plato.stanford.edu/archives/sum2016/entries/information-biological/>.

Gomez, J.C. 1994. 'Mutual awareness in primate communication: a Gricean approach', in S.T. Parker, R.W. Mitchell & M.L. Boccia (eds), *Self-awareness in animals and humans: developmental perspectives*, Cambridge University Press: New York: 61-80.

Grafen, A. 1990. 'Biological signals as handicaps', *Journal of Theoretical Biology*, vol. 144, no. 4: 517-546.

Grandy, R.E. & Warner, R. 2017. 'Paul Grice', *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), E.N. Zalta (ed), URL = <https://plato.stanford.edu/archives/win2017/entries/grice/>.

Gredebäck, G., Astor, K. & Fawcett, C. 2018. 'Gaze following is not dependant on ostensive cues: a critical test of natural pedagogy', *Child Development*, vol. 89, no. 6: 2091-2098.

Grice, H.P. 1957. 'Meaning', *The Philosophical Review*, vol. 66, no. 3: 377-388.

Grosse, G., Behne, T., Carpenter, M. & Tomasello, M. 2010. 'Infants communicate in order to be understood', *Developmental Psychology*, vol. 46, no. 6: 1710-1722.

Hauser, M.D. 1996. *The evolution of communication*, MIT Press: Cambridge, MA.

Hare, B., Call, J. & Tomasello, M. 2001. 'Do chimpanzees know what conspecifics know?', *Animal Behaviour*, vol. 61, no. 1: 139-151.

Harms, W.F. 2004. 'Primitive content, translation, and the emergence of meaning in animal communication', in D.K. Oller & U. Griebel (eds), *Evolution of communication systems: a comparative approach*, MIT Press: Cambridge, MA.

Heyes, C. 2014. 'Submentalzing: I am not really reading your mind', *Perspectives on Psychological Science*, vol. 9, no. 2: 131-143.

Heyes, C. 2015. 'Animal mindreading: what's the problem?', *Psychonomic Bulletin & Review*, vol. 22, no. 2: 313-327.

Heyes, C. 2016. 'Born pupils? Natural pedagogy and cultural pedagogy', *Perspectives on Psychological Science*, vol. 11, no. 2: 280-295.

Heyes, C. & Dickinson, A. 1995. 'Folk psychology won't go away: response to Allen and Bekoff', *Mind & Language*, vol. 10, no. 4: 329-332.

Holland, P.C. 2008. 'Cognitive versus stimulus-response theories of learning', *Learning & Behaviour*, vol. 36, no. 3: 227-241.

Hull, D. 1986. 'On human nature', in *PSA: proceedings of the biennial meeting of the philosophy of science association*, vol. 1986, no. 2: 3-13. Philosophy of Science Association.

Hurford, J.R. 2007. *The origins of meaning: language in the light of evolution*, Oxford University Press: UK.

Kaplan, D.M. & Craver, C.F. 2011. 'The explanatory force of dynamical and mathematical models in neuroscience: a mechanistic perspective', *Philosophy of Science*, vol. 78, no. 4: 601-627.

Laland, K.N., Sterelny, K., Odling-Smee, J., Hoppitt, W. & Uller, T. 2011. 'Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful?', *Science*, vol. 334, no. 6062: 1512-1516.

Leal, M. 1999. 'Honest signalling during prey-predator interactions in the lizard Anolis cristatellus', *Animal Behaviour*, vol. 58, no. 3: 521-526.

Lean, O.M. 2014. 'Getting the most out of Shannon information', *Biology & Philosophy*, vol. 29, no. 3: 395-413.

Lewis, D. 1969. *Conventions*, Harvard University Press: Cambridge, MA.

Lewis, S.M. & Cratsley, C.K. 2008. 'Flash signal evolution, mate choice, and predation in fireflies', *Annual Review of Entomology*, vol. 53: 293-321.

Liebal, K., Behne, T., Carpenter, M. & Tomasello, M. 2009. 'Infants use shared experience to interpret pointing gestures', *Developmental Science*, vol. 12, no. 2: 264-271.

Liebal, K., Call, J., Tomasello, M. & Pika, S. 2004. 'To move or not to move: how apes adjust to the attentional states of others', *Interaction Studies*, vol. 5, no. 2: 199-219.

Manser, M.B. 2013. 'Semantic communication in vervet monkeys and other animals', *Animal Behaviour*, vol. 86, no. 3: 491-496.

Laurence, S. & Margolis, E. 'Concepts and cognitive science', in E. Margolis & S. Laurence (eds), *Concepts: core readings*, MIT Press: Cambridge, MA: 3-81.

Marr, D. 1982. *Vision*: MIT Press, Cambridge, MA.

Matthews, R. 2007. *The measure of mind: propositional attitudes and their attribution*, Oxford University Press: UK.

Maynard-Smith, J. & Harper, D.G. 1995. 'Animal signals: models and terminology', *Journal of Theoretical Biology*, vol. 177, no. 3: 305-311.

Maynard-Smith, J. & Harper, D.G. 2003. *Animal signals*, Oxford University Press: UK.

Mayr, E. 1961. 'Cause and effect in biology', *Science*, vol. 134: 1501-1506.

Meunier, H. 2017. 'Do monkeys have a theory of mind? How to answer the question?', *Neuroscience & Biobehavioral Reviews*, vol. 82: 110-123.

Miller, S.A. 2009. 'Children's understanding of second-order mental states', *Psychological Bulletin*, vol. 135, no. 5: 749-772.

Millikan R.G. 2004. *Varieties of meaning*, MIT Press: Cambridge, MA.

Millikan, R.G. 2007. 'An input condition for teleosemantics? Reply to Shea (and Godfrey-Smith)', *Philosophy and Phenomenological Research*, vol. 75, no. 2: 436-455.

Moore R. 2016. 'Meaning and ostension in great ape gestural communication', *Animal Cognition*, vol. 19, no. 1: 223-231.

Moore, R. 2017. 'Gricean communication and cognitive development', *Philosophical Quarterly*, vol. 67: 303-326.

Moore, R., Call, J. & Tomasello, M. 2015. 'Production and comprehension of gestures between orang-utans (*Pongo pygmaeus*) in a referential communication game', *PLoS One*, vol. 10, no. 6: e0129726.

Moore, R., Liebal, K. & Tomasello, M. 2013. 'Three-year-olds understand communicative intentions without language, gestures, or gaze', *Interaction Studies*, vol. 14, no. 1: 62-80.

Neander, K. 2013. 'Towards an informational teleosemantics', in D. Ryder, Kingsbury, J. & Williford, K. (eds), *Millikan and her critics*: Wiley-Blackwell: UK.

Neander, K. 2017. *A mark of the mental: in defense of informational teleosemantics*, MIT Press: Cambridge, MA.

Nolan, D. 2005. *David Lewis*, Routledge: London.

O'Gradey, C., Kleisch, C., Smith, K. & Scott-Phillips, T. 2015. 'The ease and extent of recursive mindreading, across implicit and explicit tasks, *Evolution & Human Behavior*, vol. 46, no. 4: 313-322.

Onishi, K.H. & Baillargeon, R. 2005. 'Do 15-month-old infants understand false beliefs?', *Science*, vol. 308, no. 5719: 255-258.

Owren, M.J. & Rendall, D. 1997. 'An affect-conditioning model of nonhuman primate vocal signaling', in D.H. Owings, M.D. Beecher & Thompson, N.S. (eds), *Perspectives in ethology*, Springer: Boston, MA: 299-346.

Owren, M.J., Rendall, D. & Ryan, M.J. 2010. 'Redefining animal signaling: influence verses information in communication', *Biology & Philosophy*, vol. 25, no. 5: 755-780.

Pereira, M.E. & Macedonia, J.M. 1991. 'Ringtailed lemur anti-predator calls denote predator class, not response urgency', *Animal Behaviour*, vol. 41, no. 1: 543-544.

Piccinini, G. & Craver, C. 2011. 'Integrating psychology and neuroscience: functional analyses as mechanism sketches', *Synthese*, vol. 183, no. 3: 283-311.

Piccinini, G. & Scarantino, A. 2010. 'Computation vs information processing: why their difference matters to cognitive science', *Studies in History and Philosophy of Science Part A,* vol. 41, no. 3: 237-246.

Ramsey, W.M. 2007. *Representation reconsidered*, Cambridge University Press: Cambridge.

Reddy, M. 1979. 'The conduit metaphor: a case of frame conflict in out language about language', in A. Ortony (ed), *Metaphor and thought*, Cambridge University Press: Cambridge.

Rendall, D. & Owren, M.J. 2013. 'Communication without meaning or information: abandoning language-based and informational constructs in animal communication theory', in U.E. Stegmann (ed), *Animal communication theory: information and influence*, Cambridge University Press: Cambridge: 151-188.

Rendall, D. Owren, M.J. & Ryan, M.J. 2009. 'What do animal signals mean', *Animal Behaviour*, vol. 78, no. 2: 233-240.

Rescorla, R.A. 1988. 'Pavlovian conditioning: it's not what you think it is', *American Psychologist*, vol. 43, no. 3: 151-160.

Rescorla, M. 2009. 'Cognitive maps and the language of thought', *British Journal for the Philosophy of Science*, vol. 60, no. 2: 377-407.

Scarantino, A. 2010. 'Animal communication between information and influence', *Animal Behaviour*, vol. 29, no. 6: e1-e5.

Scarantino, A. 2013. 'Animal communication as information-mediated influence', in U.E. Stegmann (ed), *Animal communication theory: information and influence*, Cambridge University Press: Cambridge: 63-87.

Scarantino, A. 2015. 'Information as a probabilistic difference maker', *Australasian Journal of Philosophy*, vol. 93, no. 3: 419-443.

Scaratino, A. & Clay, Z. 2015. 'Contextually variable signals can be functionally referential', *Animal Behaviour*, vol. 100: e1-e8.

Scholl, R. & Pigliucci, M. 'The proximate-ultimate distinction and evolutionary developmental biology: causal irrelevance versus explanatory abstraction', *Biology & Philosophy*, vol. 30, no. 5: 653-670.

Scott-Phillips, T.C. 2008. 'Defining biological communication', *Journal of Evolutionary Biology*, vol. 21, no. 2: 387-395.

Scott-Phillips, T.C. 2015. *Speaking our minds: why human communication is different and how language evolved to make it special*, Palgrave Macmillian: London.

Searcy, W.A. & Nowicki, S. 2005. *The evolution of animal communication: reliability and deception in signaling systems*, Princeton University Press: Princeton, NJ.

Seyfarth, R.M. & Cheney, D.L. 2017. 'The origin of meaning in animal signals', *Animal Behaviour*, vol. 124: 339-346.

Seyfarth, R.M. & Cheney, D.L. 2003. 'Signalers and receivers in animal communicaiton', *Annual Review of Psychology*, vol. 54: 145-173.

Seyfarth, R.M. & Cheney, D.L. 2010. 'Production, usage, and comprehension in animal vocalizations', *Brain & Language*, vol. 115, no. 1: 92-100.

Seyfarth, R.M., Cheney, D.L., Bergman, T., Fischer, J., Zuberbuhler, K. & Hammerschmidt, K. 2010. 'The central importance of information in studies of animal communication', *Animal Behaviour*, vol. 80: 3-8.

Seyfarth, R.M., Cheney, D.L. & Mahler, P. 1980. 'Vervet monkey alarm calls: semantic communication in a free-ranging primate', *Animal Behaviour*, vol. 28, no. 4: 1070-1094.

Shapiro, L.A. 2016. 'Mechanism or bust? Explanation in psychology', *British Journal for the Philosophy of Science*, vol. 68, no. 4: 1037-1059.

Shannon, C.E. 1948. 'A mathematical theory of communication', *Bell System Technical Journal*, vo. 27, no. 3: 379-423.

Shea, N. 2007. 'Consumers need information: supplementing teleosemantics with an input condition', *Philosophy and Phenomenological Research*, vol. 75, no, 2: 404-435.

Shettleworth, S.J. 2010. *Cognition, evolution, and behavior,* Oxford University Press: UK.

Skyrms, B. 2010. *Signals: evolution, learning, and information*, Oxford University Press: UK.

Sperber 2000, D. 2000. 'Metarepresentations in an evolutionary perspective', in D. Sperber (ed), *Metarepresentations: a multidisciplinary perspective*, Oxford University Press: UK.

Sperber, D. & Wilson, D. 1986. *Relevance: cognition and communication*, Blackwell: Oxford.

Sperber, D. & Wilson, D. 1987. 'Precis of relevance: communication and cognition', *Behaviour and Brain Sciences*, vol. 10, no. 4: 697-710.

Sperber, D. & Wilson, D. 2002. 'Pragmatics, modularity and mind-reading', *Mind & Language*, vol. 17, no. 1-2: 3-23.

Stanovich, K.E. 2004. *The robot's rebellion: finding meaning in the age of Darwin*, University of Chicago Press: Chicago.

Stegmann, U.E. 2005a. 'John Maynard Smith's notion of animal signals', *Biology & Philosophy*, vol. 20, no. 5: 1011-1025.

Stegmann, U.E. 2005b. 'Genetic information as instructional content', *Philosophy of Science*, vol. 72, no. 3: 425-443.

Stegmann, U.E. 2009. 'A consumer-based teleosemantics for animal signals', *Philosophy of Science*, vol. 76, no. 5: 864-875.

Stegmann, U.E. 2013. *Animal communication theory: information and influence*, Cambridge University Press: Cambridge.

Sterenly, K. 2003. *Thought in a hostile world: the evolution of human cognition*, Blackwell: Oxford.

Sterelny, K. 2016. 'Deacon's challenge: from calls to words', *Topoi*, vol. 35, no. 1: 271-282.

Sterelny, K. 2017. 'From code to speaker meaning', *Biology & Philosophy*, vol. 32, no. 6: 819-838.

Tomasello, M. 1999. *The cultural origins of human cognition*, Harvard University Press: Cambridge, MA.

Tomasello, M. 2008. *Origins of human communication,* MIT Press: Cambridge, MA.

Tomasello, M. 2014. *A natural history of human thinking*, Harvard University Press: Cambridge, MA.

von Frisch, K. 1967. *The dance language and orientation of bees*, Harvard University Press: Cambridge, MA.

Weiskopf, D.A. 2011. 'Models and mechanisms in psychological explanation', *Synthese*, vol. 183, no. 3: 313-338.

Wellman, H.M., Cross, D. & Watson, J. 2001. 'Meta-analysis of theory-of-mind development: the truth about false belief', *Child Development*, vol. 72, no. 3: 655-684.

Wheeler, B.C. & Fischer, J. 2012. 'Functionally referential signals: a promising paradigm whose time has passed', *Evolutionary Anthropology: Issues, News, and Reviews,* vo. 21, no. 5: 195-205.

Wheeler, B.C. & Fischer, J. 2015. 'The blurred boundaries of functional reference: a response to Scarantino & Clay', *Animal Behaviour*, vol. 100: e9-e13.

Wilson, E.O. 1971. *The insect societies*, Harvard University Press: Cambridge, MA.

Zahavi, A. 1975. 'Mate selection-a selection for handicap', *Journal of Theoretical Biology*, vol. 53, no. 1: 205-2014.

Zuberbuhler, K. 2005. 'The phylogenetic roots of language: evidence from primate communication and cognition', *Current Directions in Psychological Science*, vol. 14, no. 3: 126-130.

Zuberbuhler, K., Cheney, D.L. & Seyfarth, R.M. 1999. 'Conceptual semantics in a nonhuman primate', *Journal of Comparative Psychology*, vol. 113, no. 1: 33-42.